Psychosocial and socioeconomic factors in the development of cardiovascular disease: a study of causality, mediation, international variation, and prediction in predominantly Eastern European settings

by
Taavi Tillmann

A thesis submitted for the degree of Doctor of Philosophy

Central and Eastern European Health Research Group
Research Department of Epidemiology and Public Health
Institute of Epidemiology and Health Care
Faculty of Population Health Sciences
University College London (UCL)

2018

Declaration of Authorship and Acknowledgements

I, Taavi Tillmann, declare that:

1. the work presented in this thesis is my own.

2. the work presented was done wholly while in candidature for a research degree at this University.

3. I have acknowledged all main sources of help. Where the thesis is based on work done by myself jointly with others, or from other sources, then I have made clear exactly what was done by others and what I have contributed myself. Namely:

   Data Analysis

   The main Mendelian randomization analyses (described in sections 2.1.4.-2.1.6) were initially performed by myself. Some of the additional sensitivity analyses (described in sections 2.1.7.1 – 2.1.7.2 and 2.1.7.5 - 2.1.7.6) were performed by Julien Vaucher (University of Lausanne). Julien Vaucher later repeated all my MR analyses, in order to minimize the possibility of human error.

   Jack Bowden (University of Bristol) supplied R code for one of the sensitivity analyses (section 2.1.7.3). Fernando Pires Hartwig (University of Bristol) performed another sensitivity analysis (section 2.1.7.4). This was necessary since Jack and Fernando had both developed new sensitivity methods which had not yet been peer reviewed at the time that I sought to perform my analyses. Kristi Läll (University of Tartu) provided me with generic R codes that I adapted and debugged along with Kristi, for the risk prediction results presented in sections 3.4.2 and 3.4.4.3. Oliver Dukes (UCL) provided me with generic R codes that I adapted for risk prediction work presented in section 3.4.3.

Ideas and concepts

The initial idea to perform the Mendelian randomization analysis was my own. To develop this, I received advice from Michael Holmes (University of Oxford) and George Davey Smith (University of Bristol).

The HAPIEE study, which sought to investigate mediation and international analysis, was initiated by Martin Bobak (UCL) and Michael Marmot (UCL). Michael Marmot (UCL) advised about the scope of my analyses, and encouraged a bolder ambition. Hynek Pikhart (UCL) provided advice about the interpretation and presentation of statistical analyses.

The initial idea to investigate risk prediction models came up in discussions with Mika Kivimäki (UCL). I received advice from Oliver Dukes (UCL) and Giovanni Veronesi (University of Insubria) about best practice in prediction science.

Abstract

Background

Psychosocial and socioeconomic factors have previously been associated with cardiovascular disease (CVD). I ask: 1) Are these associations causal? 2) Do their pathways overlap with one another? 3) Can they account for international differences in CVD? 4) Can they improve clinical risk prediction models?

Methods

Causality between education and coronary artery disease was investigated with Mendelian randomization analyses. For mediation and international differences, data on participants aged 45–69 years from the population-based HAPIEE cohort study were analysed using Cox regressions. A novel risk prediction model was derived from this data, and external validation was performed using data from the Estonian BioBank study.

Results

1) Genetic predisposition towards longer education was associated with a reduction in coronary artery disease (Odds Ratio=0.67 [95% CI= 0.59 to 0.77], per extra 3.6 extra years of education), as well as large reductions in smoking. 2) In observational analyses, cardiovascular mortality was independently associated with *unemployment, low material amenities, depression, being single, infrequent contacts with friends* and *relatives*. These associations were of similar magnitude when comparing minimally-adjusted and fully-adjusted models. 3) International differences in CVD mortality between Russian & Central European cohorts remained unexplained after adjustment for conventional, psychosocial and socioeconomic risk factors. 4) Adding predominantly socioeconomic and psychosocial factors to risk prediction models improved their discrimination, clinical effectiveness, binary NRI and Net Benefit, in derivation and validation data.

Conclusions

Education is probably a causal risk factor in the development of coronary heart disease. At least 6 socioeconomic/psychosocial factors appear to associate with CVD along pathways that may be relatively independent of one another and the conventional CVD risk factors. Uncovering their mechanisms may suggest novel avenues for intervention. While the causes of international differences in cardiovascular mortality remain unclear, socioeconomic and psychosocial factors substantially improved the performance of cardiovascular risk prediction.

Impact Statement

The first long-term implication is that it supports policymakers in enacting population-wide policies that seek to lower exposure to life stress (incl. education, poverty, social isolation and unemployment). We already know that this would create a range of social benefits. After my research, we can now we can say with more certainty than before, that some of these policies will also prevent a large amount of heart disease. Such policy debates are relatively common. As more health professionals join these debates, it will make it more likely for these policies to be adopted more quickly in more places. Furthermore, my mediation analysis suggests that it is insufficient for us to rely exclusively on one core socioeconomic risk factor. Instead, policymakers should seek to intervene across a range of socioeconomic and psychosocial risk factors, including tackling social isolation, unemployment, low education and depression. I hope that such policy translation can occur anytime from the next 10 years (optimistic estimate), up until the next 50-100 years (pessimistic estimate).

The medium-term implication is that healthcare practitioners and public health practitioners become more active in advocating for these policies. I hope that this change occurs in the next 5-20 years. The short-term implication is that psychosocial socioeconomic factors could receive greater priority in reviews of the literature, professional guidelines (e.g. the European guidelines on cardiovascular disease prevention in clinical practice), medical textbooks, and health policies (e.g. UK Public Health Outcomes Framework and Marmot Indicators for Local Authorities). I hope that this occurs in the next 0-5 years.

The second implication is that general practitioners and national public health actors in Eastern European countries “catch up”, and start using clinical risk prediction tools in their everyday practice. This could prevent a large amount of heart disease in these high-risk areas, over the next 1-10 years. For example in Estonia, I have begun working with the Ministry for

Health, and partners in the local equivalent of the NHS, with whom I am seeking to implement such a programme from autumn 2018 onwards.

The third implication is that existing risk prediction tools in Western countries (such as QRISK) could be enhanced to make use of simple psychosocial and socioeconomic variables (employment; social isolation), in the next 5-15 years.

Finally, from a research perspective, my work has shown how a wide set of conventional, psychosocial and socioeconomic risk factors do not mediate most of the association from psychosocial and socioeconomic risk factors to CVD. This means that more research is needed to determine these mechanistic pathways. Modifying these could open new avenues with which to prevent CVD (incl. pharmaceutical and other approaches), as well as adding to the evidence base about causality.

Research outputs

Publications (enclosed in annex 1)

1) Tillmann T, Vaucher J, Okbay A, Pikhart H, Peasey A, Kubinova R, Pajak A, Tamosiunas A, Malyutina S, Hartwig FP, Fischer K, Veronesi G, Palmer T, Bowden J, Davey Smith G, Bobak M, Holmes MV. Education and coronary heart disease: Mendelian randomisation study. BMJ. 2017;358:j3542. Includes a video abstract, which I commissioned: www.bmj.com/content/358/bmj.j3542

2) Tillmann T, Pikhart H, Peasey A, Kubinova R, Pajak A, Tamosiunas A, Malyutina S, Steptoe A, Kivimäki M, Marmot M, Bobak M. Psychosocial and socioeconomic determinants of cardiovascular mortality in Eastern Europe: A multicentre prospective cohort study. PLoS medicine. 2017;14(12):e1002459.

Academic conferences:

1) "Mendelian randomization study identifies causal association between Education and Coronary Artery Disease". Poster presentation at: World Congress of Epidemiology. Tokyo, Aug 2017.

2) "Causal Link Between Education and Coronary Artery Disease: A two-sample Mendelian Randomization study." Oral presentation at: Mendelian randomization in the age of large-scale accessible genomics data. Bristol, July 2017.

3) "Causal Link Between Education and Coronary Artery Disease: A two-sample Mendelian Randomization study." Invited talk to: Polygenic Prediction and its Application in Social Science Conference. Los Angeles, April 2017. Available on: https://youtu.be/LS2xD1rdGrY

4) "6 psychosocial and socioeconomic factors independently predict CVD, but not differences between countries." Oral presentation at: European Public Health Conference. Vienna, Nov 2016.

Media:

1) UK, national broadsheet – The Telegraph
2) UK, national tabloid – The Daily Mail
3) UK, national tabloid – The Express
4) UK, local tabloid – News and Star
5) UK, national specialist - Times Educational Supplement
6) Australia, national broadsheet – The Australian
7) Italy, national broadsheet – QN
8) Italy, local broadsheet – Prealpina (“Infarto, chi ha una laurea rischia meno”).
9) Switzerland, national broadsheet – Le Temps
10) Estonia, national broadsheet – Postimees
11) Estonia, national broadsheet - Postimees
12) Estonia, national broadsheet – Eesti Päevaleht
13) Estonia, national tabloid – Õhtuleht
14) Estonia, national 7pm TV news – Kanal 2
15) Global – Estonian World
16) Global – Science Media Centre
17) Global – The Conversation
18) Global – WebMD

Table of Contents

2. METHODS

3. RESULTS

4. DISCUSSION

i. Introduction

a) Rationale for the thesis

Cardiovascular disease remains the biggest cause of death globally, as well as the biggest cause of loss to disability-adjusted life years. More specifically, coronary heart disease makes the biggest contribution of all the cardiovascular diseases, and it remains the number one cause of death globally, when compared to other more specific causes of death. Although rates of coronary heart disease have been falling for about half a century in high-income countries such as those belonging to the Group of Seven (G7), more research and implementation is needed for coronary heart disease rates to continue falling in the future. There remain vast international and socioeconomic inequalities in coronary heart disease. More research is required to understand the origins of these differences, in order to suggest solutions about closing these gaps.

Progress on cardiovascular diseases prevention has largely come from two streams: first, the discovery and modification of behavioural risk factors (such as smoking and diet). Second, the discovery and modification of biomedical risk factors (such as hypertension and raised cholesterol), combined with improvements in treating symptomatic cases. One promising domain of research is psychosocial and socioeconomic risk factors. Epidemiological studies over the past five decades have shown remarkably large and consistent associations for many of these risk factors with cardiovascular disease. Despite this, they are not yet recognized as classical cardiovascular risk factors in the conventional sense, and most doctors do not enquire about nor seek to modify psychosocial and socioeconomic risk factors. As for why, a part of this may stem from misplaced discipline-specific preconceptions. For example, psychosocial and socioeconomic risk factors may be dismissed as being too ill-defined, subjective, psychosomatic, lacking biological basis, and/or lacking strong evidence of causation. Nonetheless, many of these reasons do relate to on-going areas of uncertainty, which this thesis aims to inform.

Overall, the thesis aims to advance our understanding of the links between psychosocial/ socioeconomic factors and cardiovascular disease, in order to develop practical tools and recommendations for clinicians and public health practitioners, who wish to prevent cardiovascular disease. It gives special focus to regions of the world where rates of cardiovascular disease are highest, as well as to more vulnerable people at higher risk of cardiovascular disease within such countries.

b) Structure of the thesis

This thesis has six main chapters, which are ordered to approximately follow the structure of a medical peer-reviewed article. Most chapters are themselves subdivided into around four subsections, to reflect the four overarching aims of the thesis. Chapter 1 summarizes literature on each of these four themes, beginning with the observational associations between psychosocial/socioeconomic factors and cardiovascular outcomes. Evidence is then summarized from other study designs with enhanced causal inference. This is followed with a description of international differences in cardiovascular disease, their temporal trajectories, and dominant theories to account for these. Finally, it describes the theoretical approaches to developing risk prediction models for clinical settings, how to evaluate new models, and examples of cardiovascular prediction models in use by general practitioners [a.k.a. family doctors] across Europe. Chapter 1 concludes by identifying four gaps in the evidence, which this thesis seeks to address (causality, mediation, international differences and prediction). It lists four overarching aims, with each one matched to one of the four gaps in the literature. A theoretical causal diagram is presented as a hypothesis to unite many of these aims, followed by more specific objectives and research questions.

Chapter 2 describes the methods used. The first subchapter (“Mendelian randomization”) addresses aim nr 1, and the last subchapter (“prediction”)

addresses aim nr 4. The middle subchapter (“mediation and international differences”) addresses aims nr 2 & 3, as the methods used to achieve these two aims overlap substantially. The sources of data used to answer these questions can be divided in similar ways – data from the HAPIEE study (Health, Alcohol and Psychosocial factors In Eastern Europe) is used to answer all four aims of the thesis. Additional data sources are used for aim nr 1 (other observational data, and public Genome-Wide Association Study [GWAS] data), as well as for aim nr 4 (observational data from the Estonian BioBank).

Chapter 3 presents results, with its three subchapters matched to the three methods chapter. In this way, each subchapter of the Methods, Results and Discussion sections begins to reflect a conventional peer-reviewed article.

Chapter 4 discusses the interpretation of the results. Each of the three subchapters discuss my results in comparison to the existing literature, strengths and weaknesses of my work, as well as implications for clinicians, policymakers and for future research.

1. Literature review

1.1. Psychosocial and socioeconomic factors (and CVD)

1.1.1. Scope

Given the breadth of topics covered in this thesis, section 1.1 will not account a systematic literature review of each psychosocial and socioeconomic factor. Instead, each risk factor will be introduced with its theoretical underpinning and origin, followed by one of the first reported associations with CVD. Next, in case that a recent meta-analysis is available, this will be presented to summarize the current state of evidence, thereby omitting a detailed history of the intermediating trajectory of research. The primary outcomes of interest are twofold: first, the incidence of *fatal/nonfatal coronary heart disease (CHD)*, as this is the outcome of my analysis assessing causality. Second *CVD mortality* is also reported, as this is the outcome of my subsequent sets of analyses. In the absence of published data on CHD or CVD outcomes, evidence for and against *all-cause mortality* will be presented instead. Studies where CVD risk factors are the outcome are not mentioned here, and the issue of how any causal effects might be mediated is discussed only briefly for selected risk factors.

I will mainly present epidemiological data from prospective population-based cohort studies, in the absence of which case-control and cross-sectional data will be presented where available. Priority will be given to studies that have made efforts to exclude participants with prevalent CVD at baseline, and I will not cover studies that examined prognosis among those with prevalent CVD. Since the primary interest is in individual-level exposure, area-level exposures and ecological analyses are beyond this scope of chapter 1.1. Data from peer-reviewed journals will be presented, no books or grey literature will be considered. Where possible, I will summarize evidence for and against the hypothesis that associations are larger in the following subgroups: men (vs. women); in younger age

groups (vs. older age groups); in countries with shorter life expectancy (vs. in countries with longer life expectancy).

For each of the subsections in chapter 1.1 the overwhelming majority of research reports on data collected from Western and Northern Europe, and Northern America, which have probably given rise to thousands of articles. In case that some data are available from Russia or countries in Central and Eastern Europe (hereinafter “Eastern Europe”), special effort will be made to document this. The intention is not to suggest that a considerable proportion of research has already been conducted in Eastern European countries. Instead, the intention is to clearly acknowledge any previous efforts to investigate these phenomena in Eastern European countries, despite the fact that some of these previous Eastern European studies may have been small in size or with methodological limitations. The reader could bear in mind that if equal detail were given to studies from non Eastern European countries, then this chapter would become impractically too large.

Finally, my approach to discussing and commenting on published data on the seven risk factors presented in chapter 1.1 is intentionally selective. I assume that a comprehensive assessment of all the nuances and issues particular to this body of literature is too voluminous to replicate in each of the seven risk factors. Instead, I will capitalize on the unique features presented in the particular studies on a given risk factor, and use the opportunity to expand (for that risk factor alone) into some of the implications and complexities that such caveats may bring with them to this literature.

1.1.2. Education

1.1.2.1. Early studies

Education is one of most commonly used measures of socioeconomic status (SES). As alternatives, *occupation* is one another commonly used SES indicator often seen in studies from the UK, and *income* is another commonly used indicator often seen in studies from the USA.(1) While there is evidence that all three factors retain independent associations with health following mutual adjustment,(2) education is sometimes seen as the default factor of choice, especially for global health studies focusing on lower- and middle-income countries.(3) Moreover, countries with a history of communism previously had a social structure where social differences in income or occupational prestige were probably small.(4) Educational status is more time-stable within each individual, when compared to other measures of socioeconomic status.

While the association between poverty and/or social class and health has been described for centuries (predominantly from studies in the UK and Germany), one of the first epidemiological association between *education* and CHD incidence was reported in 1968 in the journal *Science*.(5) One year later, it was shown how this association was not attenuated following adjustment for blood pressure, smoking and body weight.(6) One of the first associations between education and mortality was published in 1973.(7)

1.1.2.2. Recent meta-analyses

Given the large quantity of papers that describe educational gradients in cardiovascular disease, it is perhaps surprising to see a limited number of meta-analyses on this topic. I am aware of three such analyses. The first focused on outcomes in acute myocardial infarction, and might be characterized as having a low threshold for including relatively heterogeneous studies.(8) The analysis included studies from both cohort and case-control designs; there were no restrictions on excluding participants with prevalent CVD; and the definition of exposure to low

education was somewhat oversimplified – while primary studies have reported education at either two, three or four categories, this meta-analysis *"Compared the lowest to the highest educational groups"*. This may be one reason behind high statistical heterogeneity (I^2=78% across 42 studies). After pooling approximately 70 000 events, the authors found how participants with lowest education have a 1.34 fold [95% Confidence Interval (CI) = 1.22 to 1.47] higher risk of CHD when compared to participants with highest education. The Population Attributable Risk Fraction was reported as 25%. There was no statistical evidence of effect modification by a range of parameters, but this may be due to insufficient studies in some of the comparison groups. Point estimates were somewhat smaller in studies that rigorously excluded those with comorbid heart disease at baseline (Relative Risk [RR] = 1.29 [95% CI = 1.16 to 1.43]) as opposed to those who did not (RR = 1.75 [95% CI = 1.14 to 2.67]), suggesting that education might be more important in prognosis, as opposed to aetiology.

A second meta-analysis focused exclusively on studies in Asia, analysing 30 reports of educational gradients in all-cause mortality (of which 10 also reported a total of 10 790 CVD deaths)(9). Again, only limited attempts were made to harmonize or transform the exposure definition between studies, potentially leading to high heterogeneity (I^2=92% for CVD mortality). This also makes it hard to compare the pooled estimate from these Asian studies (RR = 1.66 [1.23 to 2.25]) against those from Western studies reported previously. The authors also looked for effect modification (reported only for all-cause mortality), and found strong statistical support for the hypothesis that the six countries with lower levels of income inequality (Gini index < 0.3) had smaller educational gradients in heart disease (RR = 1.23 [95% CI = 1.16 to 1.35]), when compared to data from nine countries with intermediate inequality (0.3 ≤ Gini < 0.4; RR = 1.62 [95% CI = 1.46 to 1.80]), as well as data from four countries with high income inequality (Gini ≥ 0.4; RR = 1.76 [95% CI = 1.24 to 2.52]). This is generally consistent with data from Western countries, where point estimates from America (a country with high income inequality) are often

slightly larger than point estimates from Europe (a region with lower income inequality).

A third meta-analysis reported a range of cardiovascular outcomes.(10) It too had the same limitations of heterogeneous exposure definitions and high heterogeneity in associations with CVD (83% < I^2 < 99% across 16 outcomes). Since the underlying studies were similar to the first meta-analysis, the results were virtually the same for CHD incidence (RR = 1.36 [95% CI = 1.11 to 1.66]). The reported estimate for CVD mortality (reported in predominantly Western European countries) was also very similar (RR = 1.39 [95% CI = 1.26 to 1.54]. The analysis of effect modification also found larger point estimates in studies from the USA, when compared to studies from Europe or Asia.

1.1.2.3. Individual level data, comparing Western & Eastern Europe

Meta-analyses of summary level data may show substantial heterogeneity from country-specific differences in how the educational system is organized and taken up, as well as study-specific differences in how educational attainment is measured during the baseline wave, how this variable is statistically transformed in the analysis, as well as which covariates are adjusted for during the analysis. As an alternative, pooling individual-level data may offer additional insights, particularly when seeking to demonstrate and effect modification on account of country-level factors that could cluster in geographical regions. As one example, the MORGAM (MOnica Risk, Genetics, Archiving and Monograph) project has collated data from 49 cohorts in Europe, altogether with 6522 CHD events in otherwise healthy participants.(11) Although this event count is two to ten times smaller than that seen in the aforementioned meta-analyses, some of this reduction in power may be offset by smaller measurement error in the exposure. Here, education was assessed by asking participants to self-report the “number of years” they believe they spent in full-time education. The researchers transformed this into population-specific, sex-specific and birth cohort-specific tertiles. This particular

transformation presents some specific caveats. For example, if one country decided to shift the entire population's distribution of exposure (or to compress the distribution of exposure) from lower to higher education slowly over time, then an analysis such as this one will assume that the associations to health outcomes are identical in both pre- and post-intervention time periods. In other words, the analysis has assumed that education is only important in a *relative* sense, when one participant's education is compared to that of their peers, as opposed to it being important in an *absolute* sense. Nonetheless, such an analysis can still detect international differences in the size of educational gradients in health. In other words, assuming that a natural educational gradient invariably exists across time and place, these investigators could ask whether some countries created environments that nonetheless managed to dampen or mitigate the health effects of this gradient. They found some evidence that the age-adjusted Relative Risks among men were highest in Eastern European countries (2.05 [95% CI = 1.57 to 2.68]), intermediate in the UK (1.73 [1.23 to 2.43]), and lowest in Nordic and Western European countries (1.37 [1.08 to 1.66]).(12) Although the confidence intervals overlap somewhat to preclude the possibility that this could not have occurred by chance alone, these are nonetheless consistent with the wider literature that supports the following theory: that countries with higher levels of income inequality are less able to protect their most vulnerable people (as measured by relative educational status) from experiencing higher Relative Risks of CVD.

Even if the Relative Risk between education and CVD were constant between different countries, then it's important to observe what happens when some countries have higher baseline risks of CVD than others. For example, Chapter 1.3 will detail how rates of CVD are higher in Eastern European countries when compared to Western European countries. Seeking to reduce health inequalities in one Eastern European country would accordingly prevent more disease and create larger net benefits to population health, when compared to an equivalent intervention in a similarly sized country in Western Europe. If indeed the Relative Risks are

larger in Eastern European countries, then this creates an additional reason to focus efforts at such high-need areas.

To demonstrate this, the MORGAM investigators have also presented the Slope Index of Inequality (SII). This measure presents the *absolute* difference in heart disease between least and most educated tertiles, and is sensitive to differences in baseline rates between country cohorts. Their analysis reports how in men, the SII of CHD mortality was much higher in Russia, Lithuania and Poland (321 [95% CI = 150 to 483]) when compared to Germany, France and Italy (44 [95% CI = -6 to 89]). A similar but weaker pattern was seen for women (72 [95% CI = 1 to 134]; and 31 [95% CI = 5 to 56], respectively). The conventional cardiovascular risk factors accounted for just one third of the association between education and incidence of CHD.

Another methodological approach is to examine mortality certificates and link these to educational attainment described in other registries, after which case-control analyses could be conducted. Mackenbach *et al.* did this systematically for 22 European countries for deaths occurring near the year 2000.(13) They looked at over 3 million deaths, of which approximately half occurred in Central European countries, and a further 140 000 occurred in the Baltic States. This made the study sufficiently powered to examine, for the first time, whether the higher baseline rates of disease seen in the Baltics may be linked to larger and steeper socioeconomic gradients in health, as measured by absolute and relative metrics, respectively.

The study findings were consistent with both of these hypotheses. In absolute terms, the age-standardized mortality rate was about 40% higher in the Baltic States, when compared to Central European countries (Slovenia, Hungary, Czech Republic, Poland). The absolute educational gradient (measured by SII) was larger in the Baltics, being about 20% larger for all-cause mortality and 20% larger for CVD mortality. This gradient was larger for deaths from injury, alcohol and causes strictly

amenable to healthcare. However, deaths from these three causes are altogether less common than deaths from CVD. Although these three other causes of death serve useful indicators of differential underlying processes, in themselves they are unlikely to explain international differences in life expectancy in this region.

The Relative Index of Inequality (RII) is similar to conventional Relative Risk or Odds Ratio metrics, but is additionally able to account for international differences in the prevalence of a given category of education and differential category widths between countries. Results for male mortality are shown in figure 1. This shows strong statistical evidence in favour of the hypothesis that the association between education and mortality is much larger in Central/Eastern Europe, when compared to Western Europe, and that this is not only a reflection of higher baseline rates in this region, but is rather a reflection of higher *relative* rates in this region.

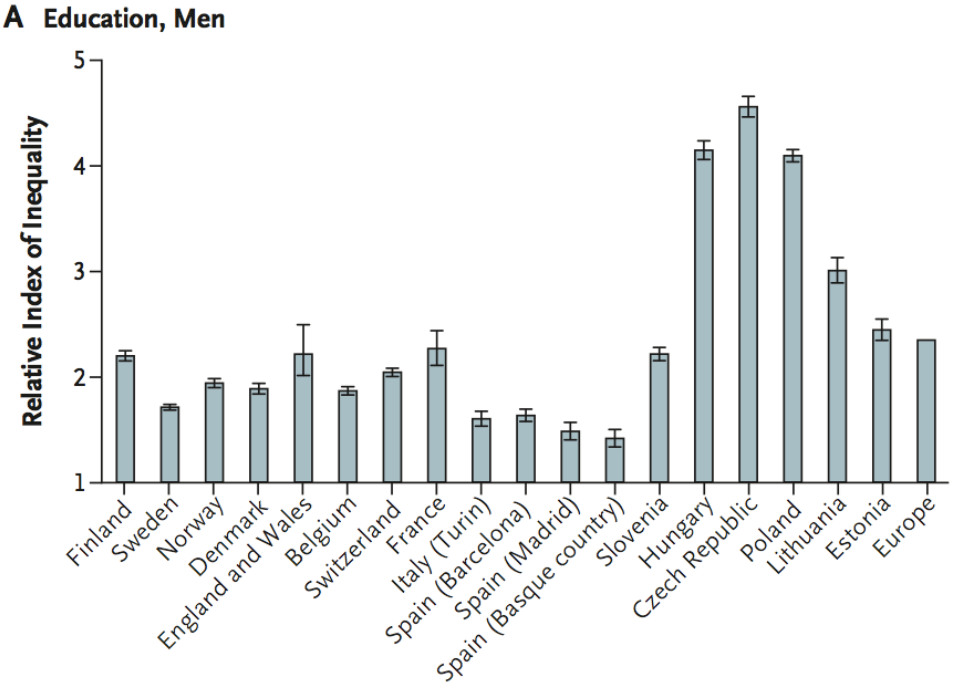


Figure 1. Relative index of inequality (RII), showing the association between low education and all-cause mortality.
High educational attainment is the reference category with RII of 1.0. Reproduced from (13).

Bobak et al. developed and applied a novel approach to collecting case-control data, whereby data on participants who had died was recalled by their spouses in one study,(14) and siblings in another study.(15) In both cases, educational attainment was associated with all-cause mortality, even after adjustment for marital status, smoking and alcohol.

The handful of Eastern European studies discussed above did not collect data about conventional cardiovascular risk factors, such as blood pressure and cholesterol. These could plausibly serve to mediate the risk from socioeconomic status to CVD (as has been investigated by many studies from non-Eastern European countries). The first study to measure these in an Eastern European country was an extension of the WHO MONICA cohort in Novosibirsk, Russia.(16) In this study, education was associated with all-cause mortality. These associations attenuated to a small degree following adjustment for the conventional cardiovascular risk factors, but no further following additional adjustment for marital status.

Previous analyses of data from the Health, Alcohol and Psychosocial factors In Eastern Europe (HAPIEE) study have also shown that the magnitude of association between education and all-cause mortality did not differ substantially between the four countries of the study (Russia, Lithuania, Poland, Czech Republic).(2) These findings are consistent with earlier prospective studies conducted in Russia.(16-18) Altogether, the current body of evidence suggest that the relative association between education and mortality is likely to be larger in Eastern European countries, but it is probably unlikely to vary within Eastern European countries.

1.1.2.4. Possible mechanisms

The potential mechanisms that might mediate the association between education and CHD, if indeed some of this is causal, have received considerable study. Again, this is a huge body of literature and I will not attempt to cover this comprehensively. As a brief summary, traditional observational associations have estimated that the association between

education and CHD attenuates by around 30-45% after statistical adjustment for health behaviours and conventional cardiovascular risk factors (including smoking, blood pressure and cholesterol).(8) However, these inferences have a wide margin of error, which can come from primarily two sources. First, models may be mis-specified, for example if confounders of the mediator-outcome association are omitted. This can bias the quantity mediated in either direction.(19) Second, any measurement error in the mediator would bias the amount mediated down towards zero.(20) Indeed, it has been shown that by measuring the mediators at two time points and modelling these as time-dependent covariates, this can further increase the amount of attenuation to nearly 100% for all-cause mortality. However, even after such a correction, the fraction mediated for CVD mortality increased only from 29% to 45%. Other studies have investigated the potential for psychosocial factors, inflammatory markers and/or healthcare behaviours as possible mediators. To date, the available evidence base suggests that these factors may account for a small part of the relationship between socioeconomic factors and CVD. It is possible that novel mediators remain to be discovered.

1.1.3. Material deprivation

1.1.3.1. Scope

Material deprivation has been defined generally as *the inability for individuals or households to [potentially be able to] afford those goods and activities that are typical in a society at a given point in time, irrespective of people's preferences to [and actual consumption of] these items*.(21) By definition, this makes comparing findings across time and place difficult, since the questionnaire themselves must be tailored to each time/place context. Typical items in such questionnaires involve a mixture of consumables (e.g. heating, utilities), more durable household assets (e.g. ownership of house, car, TV, washing machine), and some dimension of cash flow stability and/or reserves with which to prevent interruptions to what we might consider to be fundamental expenditure (e.g. recently foregone food/clothing due to unexpected shortfall in income, or due to unexpected expenditure).

Some investigators have used a single-item question as a proxy for material deprivation (e.g. car ownership, house ownership or gravestone height).(22) Others have used anthropometric short stature/height as a proxy for early childhood stunting from presumably nutritional deficiency stemming from material deprivation. However, as societies develop and stunting becomes less common, this renders the inter-individual variation of height under increasing genetic influence. Furthermore, height is socially patterned, and the genetic determinants of height appear to also associate with heart disease.(23) These nuances present a range of mechanisms by which height may be a problematic proxy of material deprivation. Perhaps for these reasons, it is less common in inventories used in recent decades of research. I will focus on studies where material deprivation is calculated as a composite, derived from more than one underlying item.

Literature searches for *"material deprivation"* primarily result in studies where exposure is operationalized at the area level. Further to the features already identified above, some of these measures additionally

incorporate area-level items on employment and social isolation. This suggests that literature on area-level material deprivation is sometimes trying to describe a higher-order construct, which can be thought of as an average of multiple psychosocial and socioeconomic factors aggregated across multiple individuals. Perhaps, these analyses are sometimes presented since such area-level data is more readily accessible to researchers. Since the data used in this thesis uses only individual-level data, I will refrain from summarizing this voluminous area-level literature. Instead, I will summarize the comparatively smaller literature where material deprivation has been measured at the individual level, and additionally where employment and social status are not part of the exposure definition (since these two are addressed in separate sub-chapters). Measures of income probably overlap substantially with measures of material deprivation, but are omitted from this thesis to facilitate a focused review.

I found surprisingly few studies that associate individual-level material deprivation with a mortality or cardiovascular outcome. Few of these are from upper-middle or high-income countries, perhaps since the pace of material growth has overtaken the pace of inventing and calibrating new questionnaires. Theoretically, the advertisement industry should supply plenty of inspiration for the content for new potential questionnaire items (e.g. Does your car have air conditioning/cruise control; Do you pay extra every month for specific TV channels/ online TV content? Does your mobile phone have unlimited data? How often do you eat out? How much you normally spend on a pair of shoes? How often in the past 3 years have you left your continent? Have you ever used a payday loan service?).

1.1.3.2. Relative versus absolute deprivation

There is a debate about whether socioeconomic factors, such as material deprivation, exert their influence in an absolute or relative manner (as discussed in chapter 1.1.2.3). A policy which shifts or reduces the average level of material deprivation without altering its distribution would improve

health only in the presence of causal effects from absolute deprivation, but not in the presence of causal effects from relative deprivation. It has been suggested that the effects of absolute deprivation may be more pronounced in lower income countries, while the effects of relative deprivation may be more pronounced in higher income countries.(22, 24) Although the evidence base to support this remains limited at current, if expanded then this may have important policy implications. For example, perhaps lower income countries are right to focus on increasing average material growth, while higher income countries might improve health quicker by shifting focus away from average economic growth toward reducing socioeconomic inequalities. To date, such research has focused on area-level measures of inequality, such as the GINI index of income inequality.(25) Such research could be complemented, in the future, with novel individual-level studies using self-reported perceptions of *relative* inequality (particularly for high-income countries), which appear comparatively lacking to date.

Some researchers have attempted to divide material deprivation into absolute and relative sub-domains. This is sometimes motivated by the view that absolute deprivation pertains more to objective measures of severe poverty, while relative deprivation pertains more to subjective reports of the absence of socially-desirable more luxury items. Vandenheede *et al.* used data from the HAPIEE study, and allocated "difficulty buying food" as a proxy for *absolute* deprivation, while amenities that lean towards the luxury scale (microwave, video recorder, colour television, washing machine, dishwasher, freezer, camcorder, satellite TV, telephone and mobile phone) represented *relative* deprivation.(2) Along similar lines Perlman *at al.* created a first index of luxury durable goods (colour TV, video cassette recorder, car, washing machine, country cabin), and a second index of basic utilities/amenities (central heating, central water supply, running hot water, centrally supplied gas, and central sewage).(17) Along similar lines, Pikhart et al. created three indices to reflect ownership of "basic needs" (a proxy for absolute deprivation),

"socially oriented" and "luxury" items (as proxies for relative deprivation).(26)

1.1.3.3. Future prospects for extending and improving this construct

It appears that decisions about which items to place into which sub-domain, and how these should be named, have predominantly been done quite subjectively with limited inter-investigator agreement, theoretical basis, or empirical justification. An empirical way to improve this could be to borrow the main approach of psychometrics, and conduct a Factor Analysis (i.e. Principal Component Analysis). This could explore the covariance patterns of the underlying items, and derive higher-order facets that are less collinear with each other. This bottom-up approach could allow higher-order items to be empirically named and conceptualized. However, one challenge would be to derive questionnaires and higher-order facets that remain stable despite substantial cultural variation manifest across time and place in terms of desirable commodities. Perhaps with the spread of globalization, much of this variation could effectively be collapsed into a linear axis of economic development.

1.1.3.4. Associations with all-cause mortality

Bobak *et al.* used an indirect case-control technique, whereby data about participants who had died were retrospectively recalled by the participant's partner. Ownership of three luxury items (colour TV, video recorder and car) predicted all-cause mortality.(14) This was not attenuated after adjustment for education, suggesting that material deprivation may be an important risk factor over-and-above education (at least in Eastern Europe).

Perlman *et al.* found that *luxury goods* were not associated with mortality, while greater *basic amenities* were associated with a lower risk of all-cause mortality in men (HR 0.95 [95% CI = 0.91 to 0.99] per each additional item, in the age adjusted model.)(17) This association became

insignificant once education was adjusted for, suggesting that education may be the strongest socioeconomic predictor of mortality in countries like Russia, and that it's pathway of effect may potentially overlap in part with that from material deprivation.

In the same study, an alternative dichotomous measure of more extreme poverty was asked with *"Have you had to sell goods [to make ends meet]"*. This was consistently associated with mortality (HR = 1.75 [95% CI = 1.17 to 2.63]), with no attenuation following adjustment for education. This suggests that some measures, such as extreme poverty, might have independent effects that are orthogonal to conventional measures of SES. However, as this particular exposure was very rare (3.5%), this is unlikely to have large effects on populations as a whole.

Vandenheede *et al.*'s measured *absolute material deprivation* (foregone food because of financial trouble) was more prevalent in the HAPIEE study: 6%12%, 20% and 33% for men in Lithuania, Czech Republic, Poland and Russia, respectively.(2) Associations with all-cause mortality were consistently high, with hazard ratio point estimates ranging from 1.5 to 2.3 across the four countries. There appeared to be no effect modification by varying the prevalence of exposure (e.g. by stratifying data from Russia Vs. elsewhere). This supports the theory that individuals make relative comparisons to their peers, and it is this subjective feeling of relative deprivation that may be harmful.

Vandenheede *et al.*'s measure of *relative material deprivation* found weaker effects when comparing the middle group against the most privileged group (HR point estimates ranged from 1.1 to 1.4), and larger effects when comparing the middle group to the most deprived group (HR point estimates from 1.9 to 2.5). This is consistent with an explanation of non-linear effects, which has otherwise been seldom addressed in existing literature of material deprivation.

In cross-sectional studies, the association between material deprivation and self-reported health disappeared after adjustment for education in one study,(27) and perceived control in another study.(28) The authors speculated that material security may enhance general psychological well-being and control, which may subsequently improve health. However, the alternate pathway is equally plausible: that individuals with a low sense of control struggle to obtain income and material resources, particularly in challenging times such as during a transition from communism towards a deregulated economy. A third study found little attenuation of the association between various measures of material deprivation and self-reported health, after additional adjustment for perceived control.(26)

To conclude, there are comparatively few studies that investigate whether material deprivation (measured at the level of the individual not area, and using multiple items) is associated with all-cause mortality. This is evident from an absence of quantitative reviews and meta-analyses on this topic. Furthermore, I am not aware any prospective reports associating individual-level material deprivation with cardiovascular outcomes from any country in the world. The few studies to date might suggest that absolute deprivation could be a stronger predictor of all-cause mortality in countries with a high prevalence of absolute deprivation, whereby high deprivation could, in turn, stem from either lower average economic development (e.g. Russia) or higher inequality in otherwise developed countries (e.g. USA). In Western European countries where absolute deprivation is less common and baseline rates of disease lower, relative deprivation might rise to become a more important risk factor. While these dimensions appear to be partly overlapping and partly orthogonal to other measures of socioeconomic position, their effect sizes have rarely equalled that of education. It is unclear whether this is due to greater measurement error when measuring material deprivation, or whether education is actually a more important predictor of health when compared to material deprivation. There is very little literature on how any putative causal effects from deprivation to disease might be mediated.

1.1.4. Unemployment

Other measures of socioeconomic status, not covered in this literature review, might include occupation and the status and prestige associated with it. One of the extreme categorizations of this might be a binary classification of whether one is unemployed or not. This is not a socioeconomic variable in the conventional sense, particularly since episodes of unemployment tend to be transient. However, such episodes are nonetheless acutely stressful, potentially on a number of levels (e.g. the material loss of income and living conditions, the sudden change in social prestige and work-related social support, as well as psychological uncertainty, worry and anxiety about the future).

Moser *et al.* were perhaps the first to demonstrate an association between individual level unemployment and all-cause mortality in 1984.(29) A recent meta-analysis of 42 studies reported unemployment to associate strongly with all-cause mortality at HR = 1.63 [95% CI = 1.49 to 1.79]. Hazard ratios were twice as large in men (1.78) when compared to women (1.37), and especially heightened for the youngest men (age < 40 HR = 1.95; age 40-49 HR = 1.86; age > 50 HR = 1.17).(30)

Some studies have suggested that the policy context may influence the effect size, with effects being bigger in countries with fewer social protections in case of unemployment (such as the USA).(31) This is consistent with observations of slightly higher point estimates from Russia. One prospective study reported a Hazard Ratio of all-cause mortality as 1.88 [95% CI = 1.38 to 2.55](32), while another case control study found the Odds Ratio for CVD mortality to be 2.55 [95% CI = 1.44 to 5.00].(33) In the latter study, unemployment and education did not attenuate substantially after mutual adjustment for each other and marital status, while the hazard ratio for marital status attenuated to null.

However, as the evidence of effect modification by country context remains rather circumstantial, alternative explanations could also be

considered for such heterogeneity. For example, the meta-analysis cited above found that the following methodological features of the study can account for heterogeneity in the final point estimate: age- and gender-mix of the sample; follow-up time (whereby hazard ratios were smaller after the 10th year of follow-up); and choice of comparison group (i.e. employed persons Vs. general population samples, including retired persons).(30) It is possible that the studies which reported higher point estimates (from Russia and the USA) were merely confounded by some of these above factors. In comparison to this literature for all-cause mortality, literature on cardiovascular outcomes is sparse and lacking a systematic review.

One interesting question is whether it is the *total time* spent unemployed, or the *number of transition spells*, that associates most strongly with health. A large study published in 2012 compiled detailed retrospective employment histories of each participant, and regressed various unemployment constructs against subsequent self-reported MI incidence (1061 events).(34) They reported a clear dose-response effect with the *number of unemployment spells.* The Hazard Ratio was 1.22 for just one spell of unemployment, and this increased with each unemployment spell until a Hazard Ratio of 1.63 was observed for those with four or more spells of unemployment. In contrast to this, measures of the *number of years spent being unemployed* showed a non-significant trend in the opposite direction (HR 1.27 for up to one year of unemployment, 0.9 for more than five years of unemployment). If these associations denote causal effects, then this is consistent with the theory that the acute phase of transition is more harmful than sustained chronic unemployment, perhaps due to the psychological uncertainty and psycho-neuro-immunological adaptive demands that such events bring.

These data are also compatible with the theory that reverse causation is unlikely to drive the observed associations: it is more plausible for the latently ill to remain chronically unemployed without transient episodes of employment. However, in this study such presumably latently ill and chronically unemployed people did not have an increased incidence of MI.

In their discussion, the authors posit an effective comparison to other cardiovascular risk factors: although single time-period measures of current smoking and diabetes status predict MI adequately, evaluating their cumulative toll over time using time-dependent measures (such as “pack years of smoking” and “HbA_1C”, both already used in routine clinical practice) allow for more precise quantification of the strain from these factors to the cardiovascular system. It is plausible that the same analogy could apply for transitions to unemployment, where measures of cumulative historic exposure (perhaps quantified by *unemployment episodes*, as opposed to the number of years unemployed) are stronger predictors than snapshots of current exposure.

As an alternative putative confounder, frequent unemployment spells might instead be a marker of employment instability, which could be caused by poor team working ability and personality traits that might predispose towards that. For example, high neuroticism is associated with both the incidence of CVD and all-cause mortality, and could putatively cause more frequent unemployment episodes.(35) I am not aware of any study where personality and unemployment were both mutually adjusted for.

Further to this study above, there have only been a handful of prospective unemployment studies were the outcome is the incidence of CVD. Nonetheless, these studies have been of good quality, size and consistency in their findings.(36, 37) Such associations were attenuated only partly following reemployment.(38) Although the association with disease incidence appears to be largest in the immediate year following unemployment and begins to decline thereafter, a statistically significant hazard can still be detected even decades after the primary event, supporting the notion that one unemployment event may lead to permanent “scars”.(39) It is also plausible that persons who never experience unemployment, but who frequently change the colleagues, location or type of their work they do, might have higher mortality when compared to those who remain employed in more stable circumstances.

Altogether, this small but consistent evidence base around health suggests a value system which is completely at odds with the value system of mainstream economic theory. There, high labour market turnover (and more unemployment events) are typically praised for delivering greater market efficiency, while most aversion is given to persistently high unemployment (for example, by monitoring national rates of unemployment). Better quantification of the health consequences of each unemployment *event*, as well as job-to-job transitions where formal unemployment is avoided, might inform and correct the perhaps excessive appreciation currently placed on high labour market turnover in economic discourse and subsequent public and political opinion.

1.1.4.2. Mechanisms

When the meta-analyses above compared age-adjusted models to fully adjusted models, then there was no attenuation in the magnitude of associations between unemployment and CVD. This suggests that, if some part of the association between unemployment and health is causal, then health behaviours are unlikely to account for much of this pathway. This observation is particularly pertinent to Eastern European countries like Russia, where those who are unemployment tend to drink more alcohol.(40, 41) One plausible mechanistic pathway would be for upstream macroeconomic stressors (e.g. privatization) to cause unemployment and psychological stress among those left in work, both of which themselves may cause an increase in alcohol consumption among both unemployed and employed persons. Each of these four factors (privatization, unemployment, fear of being made redundant, alcohol) may precipitate premature mortality. This is supported by indirect retrospective cohort data from the PrivMort study, where privatization, unemployment and alcohol consumption were all independently associated with all-cause mortality.(42) More specifically, it might be hypothesized that a part of the effect from unemployment to mortality could be mediated via alcohol, particularly in Eastern European countries. This hypothesis has yet to receive a standalone paper, since the published papers from Eastern Europe have focused on the primary effect from privatization or alcohol,

and so a more formal mediation analysis has not been explored. However, preliminary data from PrivMort appears not to support this hypothesis. The primary association between unemployment and mortality in simpler models adjusted only for education, occupation and town-level privatization (Incidence Ratio = 1.32 [0.94-1.70] in men and 1.56 [1.02-2.37] in women) was not attenuated at all following additional adjustment for alcohol consumption and other variables (Incidence Ratio = 1.40 [1.08-1.81] in men and 1.57 [1.04-2.39] in women). Accordingly, the preliminary data from Eastern Europe suggests that alcohol consumption is unlikely to be a mediator of the primary association between unemployment and mortality. However, this topic has received very little dedicated study.

One analysis from the Great Recession suggest that in the USA (but not in Europe, perhaps on account of better unemployment protection policies), those on lower incomes were particularly vulnerable to increased depression following unemployment, while those on higher incomes did not show such an association.(43) However, in this study the absolute increase in depression symptoms over time was very small (3-5%) and hence unlikely to account for much of any putative pathway from unemployment to health. As thus, the mechanistic mediators of hazard from unemployment remain largely unknown.

To conclude, unemployment has been consistently and strongly associated with the incidence of cardiovascular disease, as well as with all-cause mortality. This association appears strongest in younger men, and when additively counting the number of transition episodes into unemployment. Known cardiovascular risk factors (including health behaviours, alcohol consumption and depression) appear to account for a near-negligible part of this association.

1.1.5. Social support

1.1.5.1. Theory

Examples of how social connections influence health date back to at least one hundred years, when Durkheim described how suicide rates vary substantially by an individual's social context.(44) During the 1970s, this concept was broadened into a generalized theory of social support.(45, 46) Various conceptions of social support have been operationalized for the study of social epidemiology, such as the *structural dimensions* of a country (e.g. culture, inequality, public policy, social change), as well as the nature of *social networks* (e.g. their size, homogeneity, geographical spread, mode of communication, and reciprocity).(47) At least theoretically, the strict interpretation of such concepts is to study them at the aggregate level. For example formal network analysis requires bidirectional knowledge about whether person 1 and person 2 both agree that they believe to be friends.(48) Such aggregate measures are relatively difficult to construct and require distinct methodological approaches, so are considered beyond the scope of this review.

Nonetheless, various individual-level constructs exist that also include the word "Social Network" in their name.(49) Detailed discussion of these nuances can be found in the textbook by Berkman and Kawachi.(47) From a theoretical perspective, individual-level constructs of "Social Network" or "Social Networked-ness" might be best categorized alongside other individual-level concepts, interchangeably named along the lines of *social ties, social relationships, or structural social support.* Most commonly, these are typically derived from self-reported levels of contact with approximately four groups of people: 1) a close partner such as a spouse (for which *marital status* is a non-intrusive proxy), 2) friends 3) family members outside the household, and 4) wider community groups (incl. trade unions, charities, church groups and professional bodies), also known as *loose ties*.

The presence of *structural social support* may influence an individual's health through various mechanisms, perhaps most importantly by how

related others provide *functional social support* to the index individual during their times of need. From the other side, of those providing social support, such participation has been called *social engagement* or *social influence*.(48) A range of constructs are available to measure these functional domains, but they are again beyond the scope of this review. Thus I will use the term *social support* to denote *structural social support* at the individual level, as measured by self-reported questionnaires.

The word *support* in the phrase *social support* might denote that the index person experiences some stressor, is in acute or chronic *need,* and following this receives social support from others, which acts to compensate. This framing suggests that, social support would provide negligible benefit to those who experience few stressors. It also excludes the notion that the index person is engaged and supplies social support to others. To me, such a conceptualization resembles an inherently individualistic view of the human condition. Namely, strong individuals who are able to look after themselves have no intrinsic need for social relations and the benefit that this might otherwise provide. Social support becomes useful only for those who are unable to cope by themselves, denoting an interaction between perceived stress or perceived low self-efficacy. When we look at marital status for example, then this view might be partly supported by the observation that divorce rates and independent living may have increased over the past decades alongside increments to average material wealth. However, against this view argues the observation that even those individuals of high privilege, wealth and/or self-efficacy tend to still form pair-bonds, suggesting that the tendency to form social relationships may be more fundamental (perhaps partly innate) whose purpose may stretch beyond merely providing and receiving functional aid for those in need.

When these concepts were first mentioned in the 1970s, it was frequently assumed that social support *buffers* and protects against the harm from life stressors.(50) However, there has been very little empirical support for this notion of interactive effects. Nonetheless these early thoughts around

interactions may have influenced the thinking of Karasek, who proposed that job stress emerges as an interactive function of high demands and low control at work.(51) Job stress itself is beyond the scope of this review. However, some studies have investigated whether job stress interacts with limited social support as provided by co-workers, and found some data to support this.(52) Although such observations have led to a range of proposed theories and models, it should be noted that these observations have not always replicated.(53, 54)

If such proposed interactive effects do not exist, then it is plausible that social support may exert a homogenous effect, at all levels of life stress.(55) It could be posited that humans are intrinsically social creatures who require and benefit from social relationships, regardless of their level of self-efficacy or perceived stress. As thus, social relationships may exert their benefit not only via social support *received* from others, but also from the action of *providing* social support to others. Taking a more long-term view of human history and the human condition across a wide range of cultures, the transactional model described in the previous paragraphs (where social support interacts with stress) might merely be one of the manifestations of the pervasive influence of individualistic, transactional, non-committed and business-like manner in which economic concepts (such as utility self-maximization) have permeated our latent cultural norms and also social epidemiology. An alternative view of the underlying human condition, one often favoured by anthropologists, is that the mere act of interacting socially (regardless of any transactional need to do so, or functional utilities exchanged therein) reinforces the concept of belonging to some entity greater than the self, thereby enhancing subjective well-being.(56) If true, this might denote that the commonly used phrase "*social support*" might be best replaced with something broader, such as social *ties, relations, connections* or *existence.* This suggestion withstanding, as the term *social support* is relatively common in the existing epidemiological literature, I too will use the term "social support" in most instances hereafter.

The various mechanisms by which social support could influence health could perhaps be grouped into three broad headings, although again the evidence base behind this is extremely limited. Best understood are *health behaviours*, perhaps as these are most commonly measured. A second putative pathway is *cognitive changes*, particularly changes to subjective well-being and depression. Finally, adjustments for the above two have not always attenuated the relationships between social connections and health. While this might well be simply a reflection of measurement error in the mediators, a third set of unknown pathways have also been proposed that might be called *physical-biochemical*. This incorporates various psychoneuroendocrine alterations to the Hypothalamus-Pituitary Axis (possibly orchestrated by cortisol), subsequent inflammation, and cardiovascular reactivity. These novel biomarkers have been incorporated with existing cardiovascular biomarkers (e.g. blood pressure, BMI, glucose/HBA_1C/diabetes) in a multifaceted measure called *allostatic load*.(57) Such psychoneuroendocrine pathways have been proposed to mediate the health hazard of not only social support, but also for many other psychosocial and socioeconomic risk factors.

1.1.5.2. Links to CVD

First, I summarize data from prospective cohort studies on cardiovascular mortality, non-fatal cardiovascular incidence, and all-cause mortality. Outcomes focusing exclusively on angina are not considered, as both exposure and outcome would be prone to self-report bias. The first study in 1979 found structural social support to predict mortality from all-causes (The reported Relative Risk, across various age and sex strata, ranged from 2 to 3 with no confidence interval presented).(49) Since then, 147 prospective studies contributed data towards a broad meta-analysis.(58) This found social support to consistently predict all-cause mortality, regardless whether this was measured structurally (OR = 1.57 [95% CI = 1.46 to 1.70]), functionally (OR = 1.46 [95% CI = 1.28 to 1.66]), or with a combination of the two (OR = 1.44 [1.32 to 1.58]). Strongest effect sizes were observed when structural social support was measured in a multifaceted manner, combining marital status, friends, family and

community dimensions (OR of the point estimate = 1.91). These associations were only partly attenuated following adjustment for health behaviours, implicating the possibility that cognitive or biological pathways may additionally be involved. There was no evidence of publication bias, but moderate to large heterogeneity among the studies. Just 12 studies reported outcomes for CVD mortality, with a suggestion that associations may be slightly larger for CVD mortality, in comparison to all-cause mortality. Virtually all of these studies were conducted in Western Europe, Scandinavia or Northern America.

This broad review merged together studies that included, as well as studies that which excluded participants with disease at baseline. Some standalone studies have suggested that social support may be more important in prognostic studies of those with pre-existing disease.(59, 60) This concept has been echoed by some reviews on this topic.(61, 62). However, given the scarcity of studies that have excluded participants with disease from baseline, there is little statistical evidence to rule out the possibility that this difference may have occurred by chance alone. Virtually all studies report that men receive more protection from social support when compared to women,(63) possibly because social support correlates with socioeconomic status only in men.(64) Women tend to have larger associations with health outcomes for indices of *marital quality* as opposed to men for whom simple marital status has larger associations.(65) This is just one example of the large volume of research that has been conducted in Western countries around marital status and health. Further nuances like these are beyond the scope of this review.

One of the largest studies, where prevalent cases were excluded from the baseline sample that were followed up for nonfatal CVD, was conducted in the USA. The study was plausibly designed to detect the largest associations, by sampling men not women, and by using a multifaceted instrument of structural support. At 4 years of follow-up, social support was associated with the incidence of non-fatal stroke (91 events) but not with non-fatal MI (275 events).(66) This did not change at 10 years of

follow-up (618 non-fatal MI events).(67) However, this study sampled only health professionals, which is not particularly representative of the general population. On the one hand, this group of participants might be relatively better protected from stressors that threaten income, job security or status, suggesting that they may have low need for social supports (according to a transactional, buffering model as discussed in section 1.1.5.1). On the other hand, health professionals tend to work in a more reactive and acute setting, where witnessing the suffering of others might provoke more stressful workplace experiences than that seen in the general population. As a second weakness, it is possible that those with least social support would incur a longer delay before presenting for medical attention following a non-fatal MI, leading to differential misclassification of the outcome and bias towards the null. Another study tried to overcome this limitation by using carotid ultrasound data, to obtain objective measures of disease. They recruited women, and found marital quality to not associate with future subclinical atherosclerosis.(68) Finally, one Swedish study found the unadjusted Relative Risk of low social support to be 1.53 (95% CI = 1.02 to 2.28), in associating with a composite of fatal or nonfatal MI.(69) This effect size rose to 2.40 (95%CI = 1.36 to 4.25) when restricting the sample to men whose behavioural response was characterized as maladaptive to a stressful test. This suggests that true interactions, between social support and other personality or stress characteristics, might exist for cardiovascular outcomes. However, the latter example was based on an analysis of a subgroup with just 32 events, making chance and publication bias an equally plausible explanation.

In Eastern Europe, analysis of routine WHO mortality data does not allow for the adjustment for confounders and the exclusion of prevalent cases from baseline. Nonetheless, using this data single marital status has been a strong predictor of CVD mortality. The magnitude of the point estimates appears to have increased following the breakup of the Soviet Union.(70) Bobak *et al.*'s case-control study, using data from siblings, found marital status to associate with all-cause mortality, even after adjustment for

education and smoking.(15) This study also found a trend whereby participants who saw their sibling on a daily/almost daily basis appeared to have a lower risk of mortality, whereby this association too did not attenuate after adjustment for education and smoking. The MONICA-Novosibirsk study also found marital status to associate with all-cause mortality, even after adjustment for conventional CVD risk factors and education.(16)

Qualitative research has suggested that the family institution (and particularly resourceful wives) became a central coping mechanism during the Soviet period, which helped to counterbalance the psychological frustration and stress of being in an inflexible and unsuccessful political-economic system. In particular, the importance of family connections to gain access to goods and services that were in short supply, became increasingly acceptable in Soviet culture. This provided a material pathway via which family support could have boosted health.(71) After the transition of the 1990s, the newly embedded capitalist system has sought to replace a culture of nepotism with a culture where access to resource is determined by need, merit and/or purchasing ability. However, my own anecdotal experience in Estonia suggests that family connections are still used to allocate preferential access to taxpayer-funded healthcare resources as late as in 2017 – something which is rarely seen in Western countries like the UK. This observation might suggest that some of the associations between social support and health may not be entirely generalizable or replicated between countries from Easter and Western Europe.

To conclude, there are many operationalizations of social support, each associated with health. In general population settings, social support has consistently been shown to predict all-cause mortality and CVD mortality. Social support appears to be particularly protective after the diagnosis of a cardiovascular condition, but its role in the preceding period (among otherwise healthy participants) has been less studied. Measuring social support as a time-varying confounder may increase its strength of

association with health outcomes. In Eastern Europe, marital status has consistently been associated with CVD mortality. However, the other dimensions of social support (such as contact with friends and relatives) have been less studied. To my knowledge, no study from any country has investigated whether the negative health association from single marital status can be offset by support from other domains (such as frequent contact with friends/relatives), or whether alternatively, these facets of social support are fully independent of one another.

1.1.6. Personality traits

Objective measures of personality might be defined as interrelated tendencies in thought, feeling and behaviour. They have traditionally been studied by psychologists interested in inter-individual variation (psychometricians). Personality may be further subdivided into heritable *temperament*, and acquired *character*, although most research operationalizations ultimately merge these two. Personality is thought to be relatively stable over time, in contrast to depression which has a more fluctuating course. It is plausible that various external stimuli combine and interact with personality, to create responses (such as depression or distress) that are more or less conductive for health.

1.1.6.1. Personality types

Personality can be conceptualized categorically as *types*, or continuously as *traits*. During the 1970s and 1980s, the *Type A personality type* was found to be associated with cardiovascular disease.(72) This was picked up by the general media and much of public discourse. However, these early associations, which were done often on smaller datasets, were not subsequently replicated. Despite this initial setback, one of the three components of Type A, namely *anger and hostility*, has been consistently associated with CVD in a meta-analysis of 25 studies, to a modest degree (HR = 1.19 [95% CI = 1.05 to 1.35]).(73) However, this association attenuated substantially following full adjustment for other conventional

and psychosocial risk factors. One RCT recruited type A patients with a recent diagnosis of AMI, and modified their hostility, anger, and depression, self-efficacy and well-being. This was reported to double their survival rate [CI].(74) This leaves possible the interpretation that this particular dimension of hostility is indeed a causal and modifiable risk factor for heart disease progression, and perhaps also aetiology. It might help to consider measuring this concept not as a discrete type as in the past, but also as a continuous scale, in future studies.

1.1.6.2. Personality traits (and factors)

Personality traits are measured by asking participants to self-report their tendencies (of thought, feeling and behaviour) along multiple items, which are weighted and combined to create normally distributed traits with a particular and intuitive meaning. Many personality traits are highly correlated, and the comprehensive interrogation of these (in Factor Analysis, similar to Principal Component Analysis),(75) has allowed most psychometricians to conclude that after considering around 30 traits, these load onto five mutually independent personality factors [which I have further annotated here in brackets]: Openness to [new] experience[s], Conscientiousness [and perfectionism], Extraversion [and sociability], Agreeableness [in wishing to please others], and Neuroticism [with a pessimistic outlook].(76-78)

One individual-level meta-analysis of 3947 deaths found only *conscientiousness* to predict all-cause mortality (lowest tertile HR = 1.37, 95% CI = 1.18 to 1.58; compared to top two tertiles).(79) In a second publication, the authors pooled data from approximately one third of the original dataset, to look at associations with 423 CHD deaths.(35) As expected, a 1-SD increase in *conscientiousness* was associated with lower CHD death (HR = 0.74, 95% CI = 0.67 to 0.81). Furthermore, a second independent association was seen for *neuroticism* (HR = 1.16, 95% CI = 1.04 to 1.29), with very little attenuation after adjustment for CVD risk factors. A similar association has been reported in the UK Health

and Lifestyle Survey between *neuroticism* and CVD mortality (576 events), as well as CHD mortality (314 events), but not in a Japanese cohort with just 90 CHD deaths (80), nor in a sub cohort from the original Whitehall study.(81) As one putative mechanism, a meta-analysis of 6 studies found IL-6 to be associated with *conscientiousness*, but not with neuroticism.(82) However, it might be difficult to tease apart the sequence of cause and effect: an alternative explanation is that parental SES causes lower IL-6 (perhaps via parenting, diet, stress and/or direct genetic effects), which in turn increases (or is merely correlated non-causally) with *conscientiousness.*

One area which these two meta-analyses may have overlooked is that a smaller trait might be more specific than a larger factor, in predicting CVD (much how we saw in the example of anger/hostility being more specific than Type A personality). For example, *low perceived control* and *negative emotionality* are both part of the *Neuroticism* factor; just how *Sociability, Optimism* and *[physical- / hyper-]activity* are all parts of *Extraversion*.(83) I am not aware of a sufficiently sized analysis which has screened all 20-40 underlying factors for their association with CVD. This may be difficult to do with the existing datasets, as the largest datasets have used very brief questionnaire which are insufficient to derive accurate measures of the underlying personality traits.

1.1.6.3. Perceived control

Concepts around *perceived control* originate not from social epidemiology, but psychology (where recent interest has, to degree, been overtaken by renewed interest on the big five personality factors). Accordingly, epidemiological publications on the topic are relatively sparse. Related concept have intermittently been called either *perceived control, internal locus of control*, *self-efficacy*, *sense of mastery,* and *sense of coherence.*(84) Control can be measured at various settings, such as control over health, work, life or a combination of these. One study found that when compared to overall control as measured conventionally, stronger associations were reported for all-cause mortality if absence in

control was reported over a particular social role that appeared most important to that particular participant (e.g. control over being a functional spouse, parent, grandparent etc.).(85)

One of the earliest studies from Norway in the 1980s found low control to predict all-cause mortality in men but not in women.(86) There was also a suggestion of an association with CVD mortality: if exposure was scored from 1 to 12, then a one-unit decrease in control was associated with a HR of 1.10 [95% CI = 0.94 to 1.28; P>0.05] for CVD mortality. This study also found additional associations between measures of social participation and CVD mortality. When controlling for these, the control dimension attenuated substantially, while associations for social participation did not attenuate. Assuming equivalent measurement error and causation, this might indicate that perceived control causes social participation, which causes a lower risk of CVD. However, as this study had only 43 events, this make chance an equivalent explanation for these findings. A larger German study in the 1990s found an association between a 1-SD lower control and an increased risk of fatal and non-fatal MI (HR = 1.33 [95% CI = 1.04 to 1.72]).(87) This did not attenuate at all following adjustment for conventional CVD risk factors, other dimensions of personality and depression. Neither of these two studies excluded prevalent cases from their baseline sample.

A study from the Netherlands found control to associate with all-cause mortality, even after adjustment for education (HR = 2.19 [95% CI = 1.03 to 4.70] for comparing extreme quintiles, with approximately 150 events).(88) After 5 years of follow-up, the EPIC study found *sense of mastery* to predict CVD mortality (HR = 0.82 [95% CI = 0.72 to 0.93], adjusted for health behaviours and various personality traits including Neuroticism, in an analysis with 365 incident events), especially among those of manual social class in those with low CVD risk.(89) This was replicated after extending follow-up to 12 years, where again the association was twice as large in lower social classes.(90) Furthermore, they reported an interaction with conventional CVD risk factors, whereby

those with existing CVD risk factors showed greater associations between control and subsequent disease.

In Eastern Europe, one analysis of data from the Polish arm of the HAPIEE study found low perceived control to associate exceptionally strongly with CVD mortality, even after adjustment for education, marital status, hypertension, hypercholesterolemia, smoking, body mass index, physical activity or diabetes (HR for quartile 1 Vs. 4 = 2.68 [95% CI = 1.36 to 5.31] in men and HR = 5.18 [95% CI = 1.17 to 22.96] in women).(91) As studies from Western Europe have tended to use other transformation of their exposure variable, and furthermore given the large uncertainty of these point estimates, it is unclear whether the point estimate of the effect size may differ between Eastern and Western regions.

Finally, one of the sub-question of controls*, "I often have the feeling that I am being treated unfairly"* has been associated with incident CHD independently of traditional risk factors, job features and employment grade (HR = 1.55 [95% CI = 1.11 to 2.17]).(92) However, I am not aware of studies that have compared whether *unfairness* is a stronger predictor of health than its larger construct (*perceived control*), nor its larger construct (*Neuroticism*).

1.1.7. Depression

Clinical depression, also known as Major Depressive Disorder, is a common mental disorder, characterized by persistent sadness and a loss of interest in activities that you normally enjoy, accompanied by an inability to carry out daily activities, for at least two weeks. In addition, people with depression normally experience several of the following: a loss of energy; a change in appetite; sleeping more or less; anxiety; reduced concentration; indecisiveness; restlessness; feelings of worthlessness, guilt, or hopelessness; and thoughts of self-harm or suicide.

Clinical depression can be thought of as an extreme state along a wider axis of mood, with various states of borderline or subclinical features before one approaches the average mood of a population. Affect denotes the external manifestation of mood, such as a visible smile, while mood itself remains a subjective psychological state.

1.1.7.1 Observational epidemiology

In comparison to the previously mentioned psychosocial and socioeconomic risk factors that have been epidemiologically studied for at least 30 years if not longer, then the epidemiological study of how depression predicts physical disease is more recent. This may be due to the clinical separation of “mind diseases” from “body diseases”, making links between them appear less plausible to researchers and perhaps also funders.

One of the first prospective studies in 1994 identified how depressed affect was associated with both fatal and non-fatal CHD.(93) By 2007, a meta-analysis of 28 studies confirmed this.(94) They report exceptionally high hazard ratios (HR = 2.54 [95% CI = 2.07 to 3.10]) in those whose depression symptoms exceed clinical thresholds. The hazard ratio dropped to 1.39 (95% CI = 1.26 to 1.54) for those with subclinical depression. This observation of dose-response is consistent with the overall impression offered by other reviews. One earlier meta-analysis

suggested that clinically discrete major depressive disorder exerts an additional hazard, even when adjusted for depression symptoms.(95) This reflects the potential for a threshold effect: if symptoms cross a given threshold of severity (or they last a long time), then this may trigger biological system bifurcation, such as changes in various cognitive or neuroendocrine systems that alter their long-term set points and subsequent control via feedback loops. Such as newly programmed state may yield hazard either via cognitive-behavioural pathways (e.g. reduced help seeking) or neuroendocrine pathways (e.g. inflammation). Contrary to this biological explanation, the medical sociology explanation might posit that the act of diagnosing depression attaches iatrogenic stigma and otherwise disempowers the patient, which might manifest as greater cardiovascular risk. Other studies have found the somatic facets of depression (such as hunger, sleep and energy) to have a stronger association with cardiovascular disease than the cognitive components.(96) This offer more support for the biological interpretation, as opposed to the socio-cognitive interpretation. Finally, another study suggested that the dimension of *hopelessness* (which might correlate with depression) still exerts a hazard when depression is controlled for.(97) This is more consistent with a socio-cognitive aetiology (as opposed to biological), but might also denote the possibility of reverse causation. In Eastern Europe, the HAPIEE study found depressive symptoms to predict CVD mortality similar to reports from Western Europe. A 1 SD increase was associated with a HR of 1.20 (95% CI = 1.16 to 1.24) in men and 1.23 (95% CI = 1.12 to 1.35) in women.(98)

1.1.7.2 Experimental studies

If depression causes heart disease, then using antidepressants (among those with or without depression) could prevent heart disease. The data to support this is limited. When studying healthy participants observationally, then confounding in such data presents perhaps insurmountable problems. Users of antidepressants might have more severe depression than non-users. However, they might also be more likely to seek help and comply with medical advice (both for depression as well as for

cardiovascular risk). As a result, the use of antidepressants has not been consistently associated with cardiovascular disease using observational data.(99) I am not aware of a randomized controlled study (RCT), where antidepressants are given to people free of cardiovascular disease, to see if this can prevent the development of cardiovascular disease.

Among those participants with existing cardiovascular disease, trials of antidepressants have been noted to increase the risk of bleeding,(100) but despite this do appear to cause fewer cardiovascular events.(101) Another large trial, the ENRICHD study, did increase antidepressant use in the intervention arm from 9% to 21-28% (a mild change that was not-randomized). Antidepressants were entered as time-varying confounders, thus amalgamating both baseline and intervention effects. Nonetheless, antidepressant use was strongly associated with better cardiac outcomes, and the effect size was substantial: 33-37% reductions in cardiac events and 29-37% reductions in all-cause mortality (in age- and fully-adjusted models, respectively).(102) Robust causal inference is further challenged by the observation that depression and cardiac disease do appear to share partially overlapping genetic aetiology.(103) Despite the null findings from the randomized ENRICHD intervention and a correspondingly cautious Cochrane review,(104) a more liberal meta-analysis of 23 RCTs reported that psychological therapies do reduce mortality among those with pre-existing cardiovascular disease (OR = 0.72 [95% CI = 0.56 to 0.94)].(105) The magnitude of effect was larger in men, and when the intervention was started on only those who survived at least 2 months post-MI (OR = 0.28 [95% CI = 0.12 to 0.70), which excludes the ENRICHD design. It is possible that further refining psychological interventions at this subgroup may substantially reduce cardiac mortality among those with cardiovascular disease. However, effects were also larger when limiting follow up to only 2 years, making it possible that any causal cardiovascular benefit might be quite short lived.

1.1.7.3 Distress and Vital Exhaustion

Some scholars have defined constructs which might be thought to overlap with some personality traits, as well as with depression. One of these is psychological/psychiatric distress, as measured by the General Health Questionnaire. After an initial report in 1995 which predicted all-cause mortality,(106) a more recent report of 68 222 participants in England confirmed an association with cardiovascular disease (HR = 1.22 [95% CI = 1.14 to 1.31], per 1-SD increase in GHQ score).(107)

Another construct related to depression is Vital Exhaustion, which was developed by Appels in 1987.(108) It is defined as *excessive fatigue, feelings of demoralization, and increased irritability* and is often considered a form of adaptation to prolonged distress or burnout. Studies examining the degree of overlap with depression have yielded conflicting results. It may be that while depression contains both cognitive and somatic components, vital exhaustion focuses on the somatic component. Furthermore, it seems to me that depression is marked more by low-effort, helplessness and reduced control, while vital exhaustion is characterized instead by higher effort among those who might otherwise be more pro-active. One study associated vital exhaustion to CHD symptoms more strongly than with depression.(109) Another study from the Netherlands in 2015 did not measure depression, but found vital exhaustion to be a more important predictor of CHD, as compared to dropping systolic blood pressure from the conventional risk prediction model (increased Harrell's C-index = 0.01 [95% CI = 0.009 to 0.011], Net Reclassification Improvement = 32% [95% CI = 24 to 40%] Population Attributable Risk Fraction [PARF] = 21.1% [95% CI = 13 to 29%] in men and 27.7% [95% CI = 19 to 37%] in women).(110)

Finally, another corollary of general stress constructs is *general life satisfaction*, which one study found to associated with incident angina (perhaps due to reporting bias) but not to objective measures of CHD.(111)

1.1.7.4 Anxiety

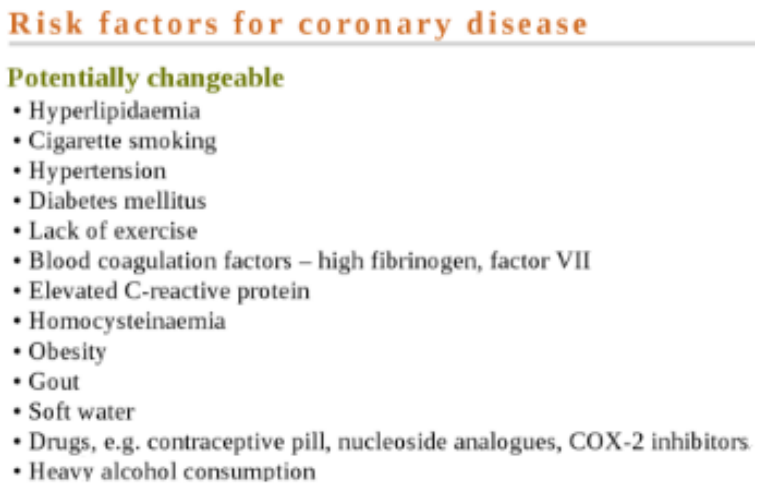
Anxiety has partial overlap with depression. A meta-analysis of 20 studies of healthy subjects found more anxious participants (with an average prevalence of around 20%) to have a higher incidence of fatal and non-fatal CHD incidence (HR = 1.26 [95% CI = 1.15 to 1.38]).(112) The magnitude of association was greater when omitting studies with composite outcomes, and instead focusing on 9 studies with exclusively fatal outcomes (RR = 1.48 [95% CI = 1.14 to 1.92]) or on 5 studies with exclusively nonfatal outcomes (RR = 1.43 [95% CI = 0.85 to 2.40]). Overall, these results are comparable those found for depression, but few of the studies on anxiety have controlled for depression. Perhaps just one study has shown how both anxiety and depression produce independent associations in univariate and multivariate models.(113) Adjustment for each other, as well as other socioeconomic variables, attenuated the univariate hazards by around one half.

1.1.7.4 Conclusion on depression

To conclude, observational studies have found a large and consistent association between depression and subsequent cardiovascular disease. One question that has not been investigated is whether it is the total time spent being with depression, or the fluctuation into and out of depression, which has the strongest association with cardiovascular disease. Causality in this whole field remains largely unexplored, due to limited experimental and mechanistic data. This might be improved by re-analysis of existing RCTs of antidepressants for long-term cardiovascular outcomes; or the application of instrumental variable analyses (where suicide rates might be the exogenous instrument, perhaps influenced by things like recession or the transition from communism to capitalism. Genetic variants for depression may also emerge in the future to sufficiently power Mendelian randomization analyses).

1.2. Is the socioeconomic-CVD association causal?

My review so far has identified how the magnitude of association, prevalence and population-attributable fraction varies considerably across psychosocial and socioeconomic risk factors. At times, these measures can even exceed that seen for conventional cardiovascular risk factors. Despite this large evidence base, there is a huge translational gap: hardly any routinely recommended clinical interventions exist (either at the individual or population level), which aim to alter psychosocial and socioeconomic risk with view to preventing heart disease. One of the manifestations of this is illustrated in figure 2.



Risk factors for coronary disease

Potentially changeable

- Hyperlipidaemia
- Cigarette smoking
- Hypertension
- Diabetes mellitus
- Lack of exercise
- Blood coagulation factors – high fibrinogen, factor VII
- Elevated C-reactive protein
- Homocysteinaemia
- Obesity
- Gout
- Soft water
- Drugs, e.g. contraceptive pill, nucleoside analogues, COX-2 inhibitors
- Heavy alcohol consumption

Figure 2. Excerpt from a popular 2017 medical school textbook, written by Kumar & Clarke (114), on modifiable risk factors for coronary heart disease.

Figure 2 lists 13 “potentially changeable” risk factors for CHD, denoting at least partial causality (and reversibility), which does not feature any psychosocial or socioeconomic risk factors. Furthermore, the evidence standard required to feature in this list appears comparatively low. For example, risk factors are included where there evidence over the last 10 years has suggested that the observational associations are probably not causal in nature (e.g. fibrinogen,(115) CRP,(116) homocysteine,(117)

gout(118-120)). I concede that some of the psychosocial factors in my earlier review are not easy to modify (e.g. personality), but the vast majority of these are theoretically modifiable. It seems to me that one argument why clinical interventions are not recommended for psychosocial or socioeconomic risk factors may be that they are perceived as being non-causal.

I will now summarize the causal evidence for and against one particular risk factor (education), since education is one factor whose causality I later study in more detail with original data. Although I will not have space to provide as much depth for each of the other factors, I will supplement my review of education with additional key references on the causality of other psychosocial/socioeconomic risk factors.

1.2.1. Education

There is a vast body of observational studies across a range of settings that report an association between education and CHD. In contrast, there have been comparatively few studies that have explicitly investigated the causality of this relationship. The existing studies on causal inference come primarily from three domains: natural experiments, sibling/twin studies, and molecular genetic data.

1.2.1.1. Natural experiments

Analyses of natural experiments have compared mortality before and after changes to compulsory schooling laws. For example, by looking at mortality rates in countries before and after the introduction of national legislation that increased minimum education, using the method of regression discontinuity analysis. In the Netherlands, such changes were associated with reductions in all-cause mortality.(121) In the UK, the largest study so far reported causal effects on improving physical activity, BMI, blood pressure, diabetes, stroke, CHD, and all-cause mortality.(122)

An extension of this design is to compare geographical areas, such as the various states within the USA. These studies initially suggested a large effect on all-cause mortality, but this effect disappeared when state-specific baseline trends were taken into account.(123, 124) In Sweden, an intervention to extend compulsory schooling throughout a 13-year transition period in a stepped-wedge design across multiple municipalities reported lower all-cause mortality in those deaths occurring after age 40 (equivalent to a hazard ratio of 0.86 (95% CI = [0.77 to 0.96]) per 3.6 years of additional education).(125)

1.2.1.2. Twins and siblings

Another source of causal inference comes from studies on monozygotic twins. Within each pair, both twins are exposed to the same set of genetic exposures (both twins are also equally exposed to some risk factors in the environment, technically known as the "shared environment"). Consequently, any difference in disease outcome between twins cannot arise from genetic effects (or factors found in their shared environment). If differences in outcome associate with differential exposure to non-shared features of the environment (such as one twin pursuing education longer than the other twin), and if the magnitude of this association is comparable to that seen in the general population, this makes less likely the possibility that the observational association is confounded by genetic (or shared environmental) factors.

Although the twin method does not eliminate the possibility of confounding from other factors in the non-shared environment, it is a design with which to eliminate the possibility of confounding from genetic factors. Twin studies initially found evidence both for and against causal effects from education to mortality and CHD incidence.(126-129)

Much of this early discrepancy could be explained by insufficient statistical power, as well as what criteria are used to judge whether the data are consistent or inconsistent with an interpretation of underlying "causal

effects". For example, analyses conducted within twin pairs often have substantially smaller sample sizes, rendering their confidence intervals wider. If this confidence interval crosses the null, then some authors and commentators might interpret this as *"no statistical evidence of causal effects"*. Although this is correct in the narrow sense of the term, such a conclusion can also be misleading, when a study is simply underpowered to definitively answer causal questions. Here, the safest interpretation is to not make an interpretation. However, if one is asked to hazard a guess, then it might be more helpful to examine whether the point estimate attenuates, when comparing results obtained from the between-twin analysis against the general population analysis.

The largest study to date from Sweden (which has twice the statistical power of the previous largest study) found strong statistical evidence for causal effects, whereby the 99% confidence intervals did not cross the null in any of the analyses.(130) Furthermore, the association between years of education and lifespan did not attenuate at all when the conventional population based analysis was compared against the between twin analysis. Hence overall, the twin literature suggests that, although only a handful of sufficiently powered studies exist, shared environmental factors (such as parenting) might not cause substantial confounding. It also suggests that confounding from genetic factors (such as genetic differences in drive, motivation, personality, or innate intellect, all of which themselves may predispose towards longer education) are unlikely to account for the observational associations between education and disease.

A parallel domain of research, using data from millions of non-identical siblings (that sometimes reached 100 times larger sample sizes than the twin studies), has also observed little attenuation of the association between education and subsequent mortality when comparing the general population analysis with the within-sibling analysis.(131, 132) As with twin studies, this also suggests that environmental and genetic factors shared by siblings are unlikely to confound the observational association seen

between education and disease. Nonetheless, the twin and sibling studies both leave open the possibility of confounding from environmental factors that are not shared between siblings.

Other twin studies have reported a causal effect from longer education to lower depression symptoms.(133) Depression might plausibly be one of the mediators, on a putative pathway from education to depression to heart disease.

1.2.1.3. Molecular genetic data

Finally, some recent studies have also looked at specific genetic variants for education. An association was found between parental longevity and genetic markers for education in their offspring.(134) However, causal directions and pleiotropy were not tested in this study.

I am aware of two reports which tried to apply the conventional Mendelian randomization (MR) approach to a socioeconomic exposure. Both tried to exploit the findings from a GWAS analysis of education published in Science in 2013, which found 3 single-nucleotide polymorphisms (SNPs) to associate with education.(135) In 2016, Nguyen *et al.* used these 3 SNPs as instruments to see these also associate with subsequent dementia.(136) Although the point estimate was suggestive of a causal effect, the 95% confidence interval crossed the null, suggesting that the study was underpowered and/or used a weak instrument. Furthermore, this study did not investigate the pleiotropy of its genetic instruments.

The other study from 2016 was by Cuellar-Partida *et al*, who were interested in myopia as the outcome.(137) Faced with the same problem of a weak instrument and limited statistical power, they took an unconventional approach towards selecting their instruments. They opted for a polygenic score that used the 17,749 “top hit” SNPs that associated weakly with education. In statistical terms, it was possible that 17,746 of these associations arose by chance alone. However, when summed together the entire polygenic score was strongly associated with the

education phenotype in multiple cohorts. This was associated with myopia, which they interpret as evidence of causal effects. The core problem with this interpretation is that many if not most of the 17 749 SNPs might plausibly have pleiotropic effects which associate with myopia via pathways that bypass education. This was not explored in their analysis. Another problem with this study is the temporal sequence of events: most people develop myopia before they complete their educational attainment. It may be that these 17 749 SNPs tag pathways that might predispose children towards spending more time towards reading and writing. It is plausible for the myopia-inducing effect of this activity to have its greatest causal effect between the ages of 0-16. This means that any extension of education, between the ages of 16-25, may or may not cause myopia. This example illustrates the problem of identifying “causal effects” in the abstract sense, since all causal effects are by nature tied down to a particular context of time, place and person.

No Mendelian randomization studies of socioeconomic exposures have investigated any other disease outcome, such as cardiovascular diseases. Furthermore, most of the other designs listed above (including natural experiments, twin and sibling designs) have reported outcomes for all-cause mortality. Few have reported cardiovascular mortality and virtually none have reported fatal/nonfatal CHD, as I plan to study.

1.2.1.4. Conclusion on the causality of education

Data from twin/sibling studies, as well as natural experiments, are on balance more compatible with an interpretation of causal effects from education to all-cause mortality. Both study designs make entirely different assumptions, which further strengthens causal inference via effective triangulation. Furthermore, the sibling design investigates the consequence of raising or lowering the population mean (also known as the *average treatment effect*). In contrast, natural experiments with compulsory schooling laws only alter exposure in the high-risk subgroup and may not influence the mean. Convergent results across both domains suggest that a causal effect could operate monotonically (perhaps even linearly) across a range of different educational exposures.

To counterbalance these direct investigations of causality, other inferences from more indirect sources are less compatible with a causal interpretation. For example, even after adjusting for parental and household factors, the presence of childhood health conditions (such diabetes, asthma, attention-deficit hyperactivity disorder, and obesity) have all predicted subsequent educational attainment, possibly via missed school.(138) This has been interpreted to denote reverse causation: that environmental or genetic predisposition towards poor health causes lower education.(47) However, this argument becomes more tenuous for cardiovascular disease, which does not manifest itself until much later in life. One could suppose a general “frailty phenotype” which may already be operating during childhood. It is possible that this frailty phenotype causes childhood illness, which causes low education. The frailty phenotype could, in parallel, also cause cardiovascular disease (figure 3, top). Although plausible, this interpretation is not compatible with the twin and natural experiment data discussed above. An alternative explanation is a simple serial one, where frailty causes childhood illness, which causes low education, which causes heart disease (figure 3, bottom).

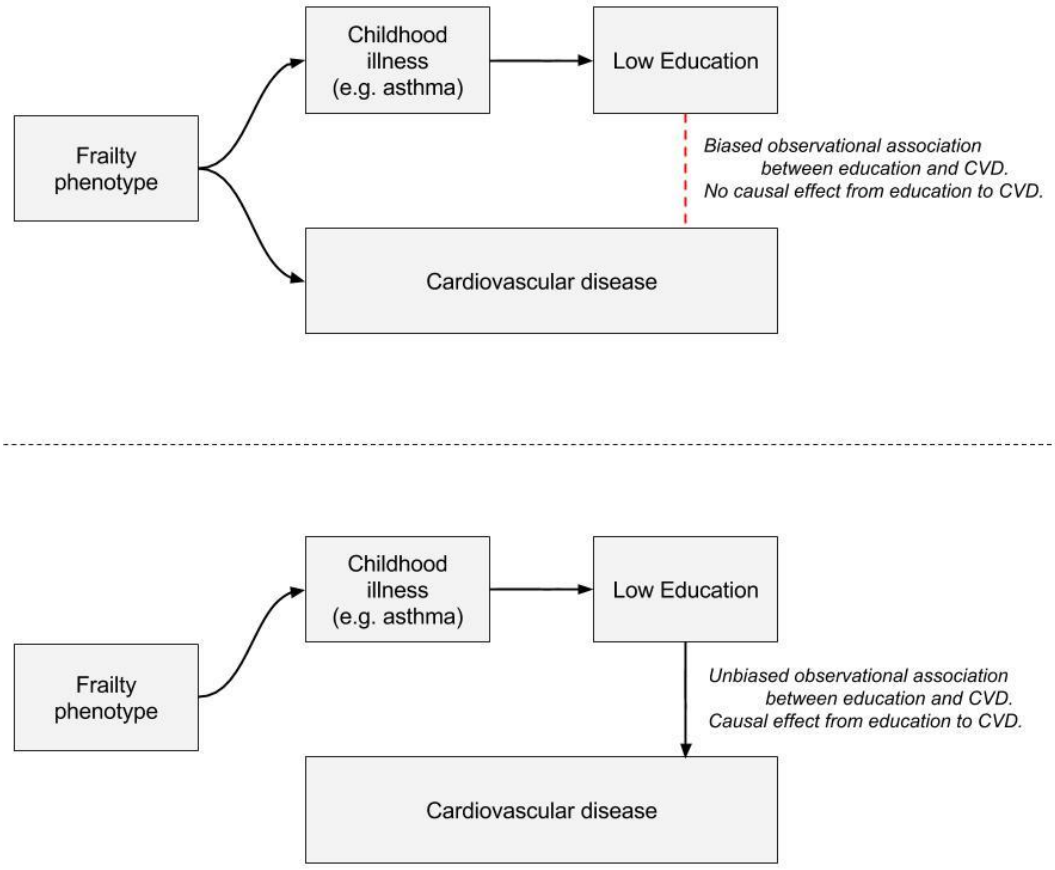


Figure 3. Two alternative indirect inferences, made from the observation that childhood illness associates with lower subsequent educational attainment.

A second indirect inference comes from the fact that educational gradients in smoking behaviour are typically established before educational attainment is completed. For example, such gradients are already seen at the age of 17, when grouping participants based on their future educational attainment.(139) One interpretation is that the pathology of smoking and subsequent illness happens independently of education. Instead, the common cause of both smoking and low education might be something such as “excess focus on the short term, at cost to long-term”, and so education is merely a non-causal marker of this (figure 4, top). However, this interpretation is also incompatible with the natural experiment data.

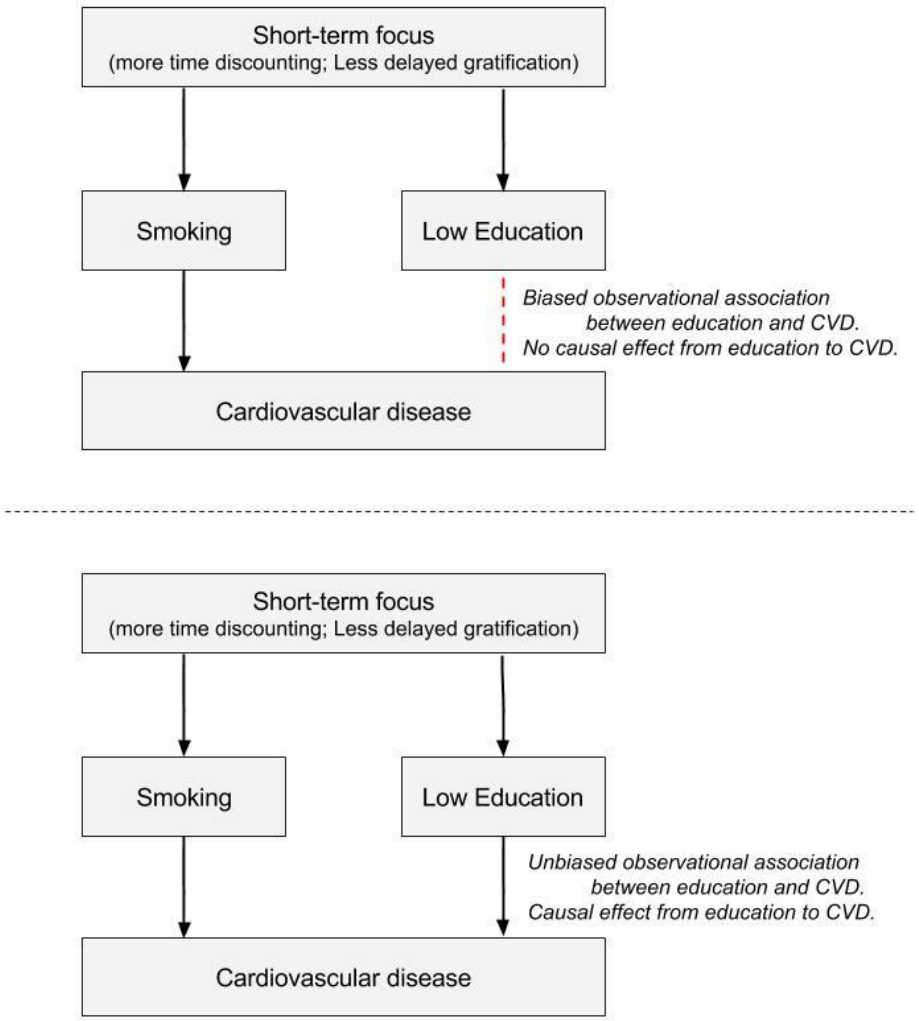


Figure 4. Two alternative indirect inferences, made from the observation that educational gradients in smoking are established before the completion of education.

Another interpretation is that education is a consequence of “short-term focus”. Nonetheless, this does not stop education from also being a cause of subsequent cardiovascular disease.

In conclusion, indirect evidence, such as the two examples discussed, offer some criticism of the model that lower education causes heart disease. The direct causal evidence, from twin/sibling studies and natural experiments are by no means definitive, but both sources suggest that a causal effect is more likely than not.

1.2.2. Unemployment

1.2.2.1. Twin studies

The twin design has also been used to study unemployment as the exposure, and all-cause mortality as the outcome. One study reported results separately for men and women.(140) My meta-analysis of their data suggests that if one twin is discordant on unemployment, then this is associated with a Hazard Ratio of mortality of 1.62 (95% CI = 1.35 to 1.89). This suggests that unmeasured genetic confounders and confounders in the shared environment are unlikely to drive the observational association between unemployment and mortality, which may indeed be causal at the individual level.

1.2.2.2. Natural experiments

Since unemployment can be a comparatively transient and dynamic state (as opposed to educational status is which is more permanent), the epidemiological study of exposure to unemployment has more explicitly discussed the possibility of reverse causation, also known as “health selection”. Here, healthier individuals could be selected into employment, while latently unhealthy individuals may find themselves more likely to be made redundant and/or having less success in their job applications. Martikainen *et al.* showed how the magnitude of association, between unemployment and subsequent mortality, varies by as much as twofold in Finland, when comparing a period high countrywide unemployment with a period of low countrywide unemployment.(141) In times of recession, when unemployment rates are higher, the strength of association with subsequent mortality was substantially lower. This is compatible with an explanation that the cohort of participants who are employed during economic grown, but face job loss during recession, are healthier than the cohort of participants who are chronically unemployed. This is compatible with a hypothesis of health selection/reverse causation. Another interpretation of this data is that length of time spent being unemployed has a monotonic dose- response causal effect on subsequent mortality:

the former cohort have simply been exposed less, while the latter cohort have been exposed more. In both cases, causal effects may operate.

The health selection/reverse causation hypothesis has also been supported by observations during the Great Recession, during which individuals with a chronic illness were twice as likely to lose their job when compared to healthy individuals.(142) Most studies of the great recession looked at mortality outcomes at the aggregate level, without access to individual level data. Such studies found that greater national unemployment rates were associated with increased suicide rates.(143) Some investigators have also reported increased unemployment to associate with *lower* national levels of cardiovascular and all-cause mortality,(144) especially among the working age population.(145) These results have been contested on methodological grounds,(146, 147) but more importantly it is inconsistent with the opposite direction of association found from epidemiological studies that use individual-level data. Indeed, this example has been held as an example of ecological/individual fallacy: that associations at the individual level of analysis may not be found in higher levels of analysis. Nonetheless, a recent synthesis looked at individual level data over a period of 18 years with varying rates of unemployment.(148) The researchers essentially looked for a binary two-by-two interaction between individual-level unemployment and area-level unemployment. The results reconciled the previously opposing results, by confirming that recessionary periods amplify the individual-level mortality hazard for those people who are made unemployed (hazard ratio of recession = 2.54, 95% CI = 1.39 to 4.65). However, for the vast majority of individuals who managed to escape being made redundant, periods of recession were associated with large reductions in mortality (hazard ratio of recession = 0.50, 95% CI = 0.31 to 0.78). Altogether this more nuanced picture of the subgroup-specific associations reconciles the findings of previously critical studies by Martikainen *et al*.

1.2.2.3. Conclusion

One way to synthesize this body of evidence, in my personal opinion, would be to conceptualize individual-level unemployment (and interventions which might lower that), as conceptually different to area-level unemployment (and interventions which might lower that). If one were to take a single unemployed person and put them into employment, then there is a reasonably strong evidence base that this is likely to improve their well-being, mental health, and all-cause mortality. Such causal effects, at the individual level, may be larger during times of recession.

As for what happens when a country lowers its national unemployment rate, then this remains less clear. Mortality from suicide will probably fall while mortality from road traffic accidents and alcohol-related causes will probably increase. Mortality from cardiovascular disease and all-cause mortality may increase, but the evidence for this is extremely limited. Based on this historical picture, it is likely that future recessions may create similar effects, at the national level. However, it also remains possible to decouple these effects using other societal factors (such as government social spending, and/or active labour market programmes). It is also possible that these associations are confounded, or modified by other factors, such as changes to disposable income, as well as popular predictions about the duration of recession and the feeling of hopefulness or hopelessness that this may bring. For example, a recession that is expected to resolve quickly may promote a positive adaptive response among those still employed, similar to the "flight or fight" response (e.g. *"drink less, drive less & savour your good fortune").* By contrast, a prolonged recession with little hope, such as that seen after the collapse of the Soviet Union, may promote a distress response (e.g. *"drink more & despair at your poor fortune"*). I posit that one mechanism through which increases to the national unemployment rate could cause reductions to all-cause mortality (driven by the large fraction of people who remain in employment) might be via the mechanism of hopefulness or "I've had better luck than my neighbour". This could create a psychological state

that could protect against depressive symptomology which may cause cardiovascular disease. Although such a theory could be tested using data similar to that used by Noelke et al, (148), I still struggle to see the real-life translation of such research: should governments seek to tolerate occasional acute recessions, in order to lower population mortality, or should they seek to reduce their occurrence, in order to protect those most vulnerable and lower health inequalities?

1.2.3. Psychosocial risk factors

A relatively small study of 2350 twins reported causal effects from divorced/never married marital status, and subsequently greater smoking behaviour and depression symptoms.(149) I am not aware of natural policy experiments that have altered marital status.

Two randomized controlled trials (RCTs) have altered depression and social support with the intention of preventing recurrent cardiovascular events and delaying mortality. The ENRICHD trial did not find an effect, while the SUPRIM trial found a benefit from using Cognitive-Behavioural Therapy.(102, 150) Both trials recruited patients with pre-existing cardiovascular disease. It may be that interventions at this stage of disease aetiology find it challenging to alter sequelae. To my knowledge, there are no RCTs that altered psychosocial or socioeconomic exposure in otherwise healthy participants with the intention of preventing the development of CVD. It is plausible that as traditional risk factors recede with period effects and effective public health intervention, this opens up the possibility that the fraction preventable from psychosocial interventions may correspondingly increase. In other words, cardiovascular disease may be becoming less common, but more psychosocial in its nature.

1.3. International differences in CVD

1.3.1. Trends in mortality across Eastern and Western Europe

Temporal fluctuations in mortality and life expectancy have been more dramatic in Eastern Europe than anywhere else in the world, when excluding periods of war and famine. Figure 5 shows most of the countries from Western Europe as coloured green, and most of the countries from Central- and Eastern-Europe as coloured red. (Of note, membership of the OECD has slightly changed since the production of this figure.) As the definition of “Central and Eastern Europe” can vary, in this definition three countries (Poland, Slovakia, Hungary) that are conventionally be thought of as being in Central/Eastern Europe, are here marked green as OECD/Western Europe instead. Nonetheless, other than for these three countries, figure 5 provides a reasonable approximation of the trends in life expectancy across these two regions.

Trends in Western Europe have been remarkably stable, showing a near linear increase in life expectancy over a 65 year period, with little change in the rank between each country. The absolute speed of progress for a country such as the UK has been around 2.2 months of additional life expectancy gained, pear each calendar year.

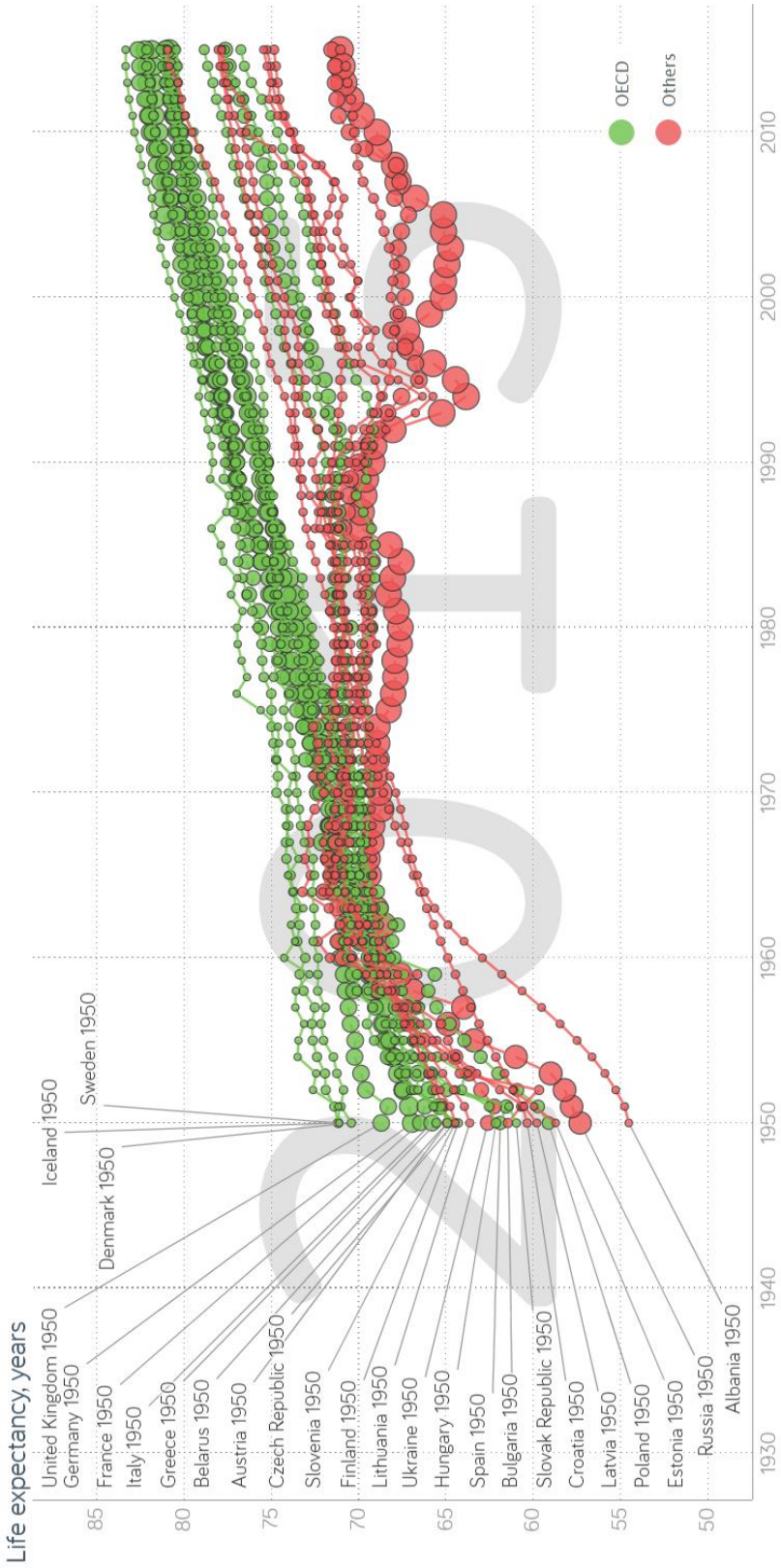


Figure 5. Trends in life expectancy (1950-2015) among selected Western and Eastern-European countries.

Visualization website: Gapminder.org

Data source: United Nations Department of Economic and Social Affairs.

Trends in Eastern Europe, however, might be characterized into four periods. Between 1950-1965, life expectancy improved dramatically. In Russia, life expectancy improved by 10 months per each calendar year. This period coincided with Nikita Khrushchev's leadership of the Soviet Union. By 1965, life expectancy in each of the three Baltic States, for example, was higher than that seen in Finland or Austria.

Between 1965-1990 improvements in life expectancy stagnated. This period coincided with the wider "Era of Stagnation" (predominantly with the USSR being under Leonid Brezhnev's leadership) when little progress was also seen in other aspects of society (such as cultural freedoms and economic growth). During this era, there was one notable temporary deviation: life expectancy increased in Russia very quickly, by 14 months per year between 1984-1986. This is largely attributable to Gorbachev's anti-alcohol legislation. However, life expectancy in the subsequent two decades fell and never regained the temporary peak it attained during 1986.

More specifically, between 1990-1994, the collapse of the Soviet Union was accompanied by the largest fall in life expectancy ever seen without war or famine. In Russia, life expectancy fell by 5.4 years (at a speed of -16 months of life expectancy, per each calendar year). Graphically, this is equivalent to going back to 1957, i.e. the loss of 37 years of intermittent progress. This decline accounts for why international inequalities in life expectancy diverged spectacularly between 1990-1994. Furthermore, inequalities within Eastern European counties diverged. Countries that were members of the Soviet Union saw large decreases to life expectancy, while other communist countries elsewhere in Central Europe did not see such a decline.

Between 1994-2015, the overall pattern is one of slow and almost parallel progress in East, Central and Western Europe, with little changes to the rank order of countries. Some smaller fluctuations are still visible. Due to the Russian Rouble crisis, between 1998-2003, life expectancy in Russia

fell by 1.5 years. At the same time, life expectancy in Estonia increased by 2.5 years (speeds of -4 months and +6 months / calendar year, respectively).

Estonia is also the ex-USSR country that has made the biggest progress against closing the mortality gap seen against the other Central European countries. In 1994 this gap with Poland was 5.4 years of life expectancy. By 2015 Estonia had overtaken Poland (with average annual increases of life expectancy at +7 and +3 months, respectively). Nonetheless, ignoring these smaller fluctuations still shows an overall pattern between 1994-2015 where improvements in countries like Russia (+4 months of life expectancy per calendar year) were slightly faster than corresponding improvements in the West (e.g. +2.5 and +2.7 months of life expectancy per year, in the UK and Spain, respectively). This slight increase in speed in Russia has not been at all sufficient to close the large international gaps in life expectancy that were created during the fall of communism. For example, were trends from these two decades extrapolated linearly, then life expectancy in Russia would overtake that in the UK only by 2095 (at a life expectancy of 98). As another perspective of the vastness of the current gap, life expectancy in Russia today (71.0 years) is only one year longer than it was in Russia 50 years ago (69.8 years). Furthermore, male life expectancy in Russia (62.0 years) is dramatically lower when looking at countries with comparable levels of GDP per capita (approximately 72 years), and lower than in the countries named in figure 6, whose economic development is also substantially lower (e.g. India and North Korea).

By further stratifying these trends by age and sex, it appears that the strongest fluctuations over time have occurred in men aged 25-64. In absolute terms, these changes were driven by changes to cardiovascular mortality, most specifically to changes in ischaemic heart disease (data not shown).

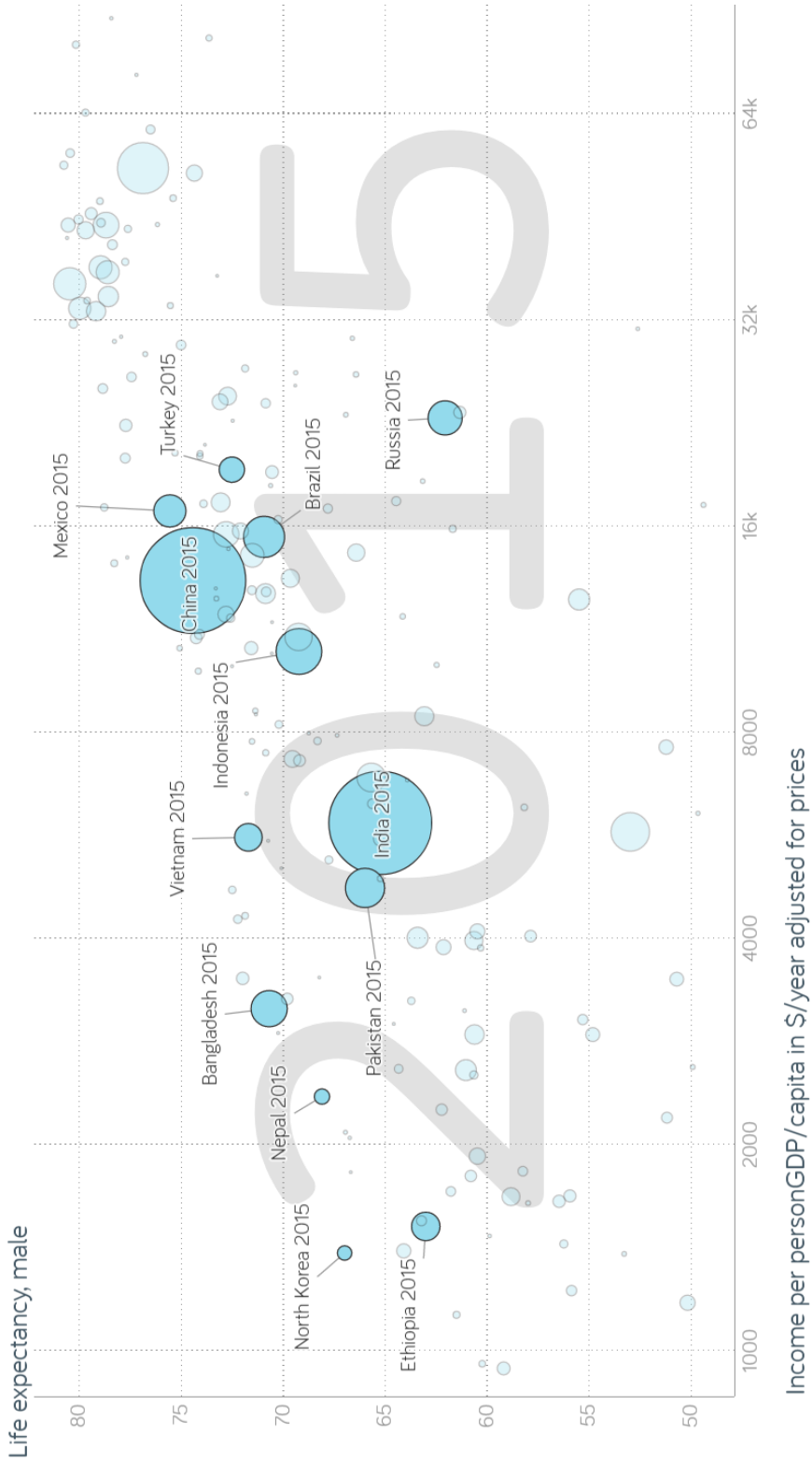


Figure 6. Scatter plot of all countries, comparing economic development against male life expectancy. *Highlighted countries have lower levels of economic growth and higher life expectancy than Russia, and have populations >25 million. Visualization website: Gapminder.org Data sources: World Bank; United Nations Department of Economic and Social Affairs.*

1.3.2. Proposed explanations

1.3.2.1. Conventional cardiovascular risk factors

The Multinational MONItoring of trends and determinants in CArdiovascular disease (MONICA) study was a detailed investigation of CHD cases from 38 countries, spanning West and East Europe between the mid-1980s to the mid-1990s. Although they did not collect detailed exposure data from the cases, they used an ecological design to look at national trends in conventional CVD risk factors from other data sources. The conventional factors were blood pressure, consumption of cigarettes, serum cholesterol, and BMI, which were assessed from 2-3 time points. This was used to construct a linear time trend in exposure, and calculate a mean annual change in exposure.(151) The authors regressed time-trends in CHD rates against time-trends in risk factors. Better fit was seen with a 4-year lag between changes to exposure and changes to CHD rates (figure 7).

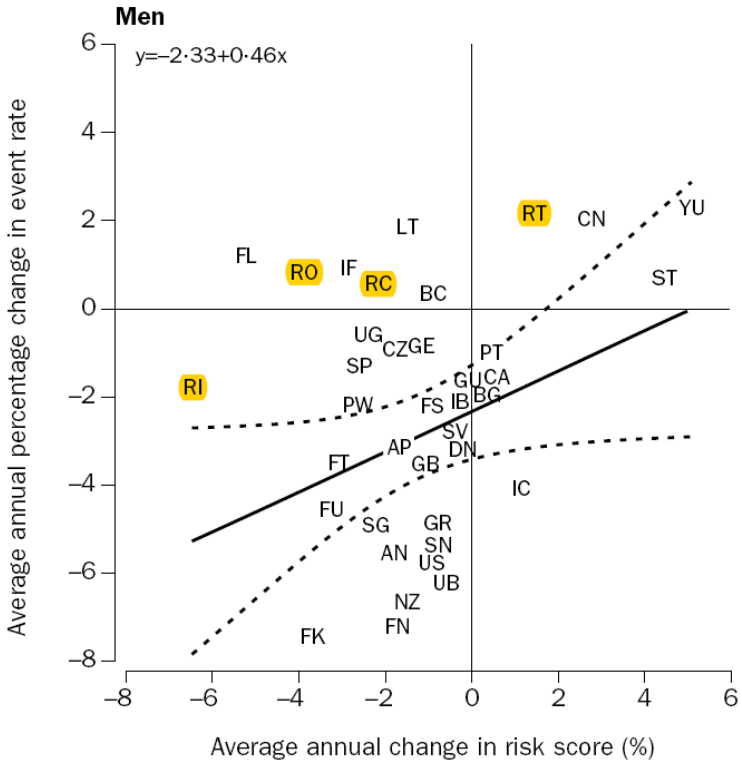


Figure 7. Lagged regression between mean changes in CVD risk factors between 1984 to 1993 (x-axis) against mean changes in CHD events between 1988 to 1993 (y-axis).
Yellow highlight denotes four studies from Russia.

In figure 7, the regression coefficient was r = 0.46 (0.01 to 0.91) This means that, assuming that temporal changes within Western and Eastern countries had similar origins, about 22% of the within-country change in CHD is attributable to within-country change in conventional risk factors. Of note, this may be an underestimate, since all measurement error in risk factors is likely to bias the correlation coefficient towards the null. There are also methodological considerations. For example, using multiple regressions for each of the four risk factors (instead of one single regression), the percentage finally explained increased to 46%, however, with substantially greater risk of overfitting and larger uncertainty (as few of the four risk factors showed conventional levels of significance).

Of note in the plot above, all of the yellow Russian samples are consistently above the trend line. The absolute annual improvement in CHD in Russia was about 4 percentage points less than might be predicted from the otherwise decline in conventional CVD risk factors. One might be tempted to exclude the Western data, and when looking only at the four Russian data points themselves, then their correlation might be informative. However, this correlation is more likely to arise from random error. For example, both the extremes (RI and RT) are in fact intervention arms, which might plausibly be expected to appear in the bottom left corner.

Altogether, although the MONICA study did tell us about the possible causes of temporal changes in CHD trends (controlling for international variation), it was not designed to account for international differences in baseline CHD, as it did not analyse individual-level data. Furthermore, it was not sufficiently powered to look at the causes of temporal variation within Eastern European countries alone. However, they were able to attribute up to half of the temporal variation in an average country (pooling data across West and East) to changes in conventional CVD risk factors. This leaves open the possibility that (assuming zero measurement error) around half of the causes of temporal variation remain to be found. This study could be improved methodologically. For example, more

sophisticated methods for capturing the total lagged effect over a wider longer tail period have been recently proposed,(152) which might increase the attributable amount. It is also possible that changes to conventional risk factors are themselves confounded by other causal confounders (for example, increased societal democracy and optimism). It is also possible that results are different in Eastern Europe, where temporal variation in outcomes was much larger. Nonetheless, the reasons as to why Eastern European countries experienced substantially worse temporal trajectories than Western countries appears to be unexplained, and probably unrelated to, changes to conventional CVD risk factors.

A separate line of enquire has been developed by Capewell, who has examined the drivers of temporal changes within one country. He has applied the IMPACT model to multiple time points, each of which should specify mortality, the prevalence of conventional CVD risk factors, as well as any other proposed explanatory factors (such as the uptake of evidence-based medical treatments). These analyses attributed the gradual secular decline in mortality seen in Western European countries to improvements in risk factors and medical treatments. Most of these models have found both to play a role, with preventative factors playing a larger role. Similar results have been observed when analysing data form Poland and the Czech Republic.(153, 154) However, neither of these two countries experienced the large drop in mortality that was observed in the ex-USSR countries between 1989-1994. To date, the IMPACT model has not been applied to ex-USSR countries.

1.3.2.2. Diet

It has been speculated that diet may be one of the main reasons why Eastern Europe has seen higher total and cardiovascular disease (CVD) mortality when compared with Western Europe.(155-157) Empirical evidence to support this has been sparse, including relatively crude cross-national comparisons,(158) time-series analyses within Poland,(159) and Global Burden of Disease models, where diet-disease associations from the West are extrapolated onto the higher disease burden seen in Eastern

Europe, to infer that diet should play a large role in Eastern Europe.(160) To the best of my knowledge, only the HAPIEE study has collected individual-level data about diet from multiple countries, to make inferences about the source of international variation.(161) In a pooled analysis across countries, a composite of healthy diet indicator had a PARFs of 14% for CVD mortality, and 3% for all-cause mortality. This suggests that diet was able to predict who is at risk *within* each country. They also reported additional analyses, where the baseline rate of CVD mortality in Czech Republic was compared against that in Russia. Mortality was 2.86 (2.23-3.67) times higher in Russia in a model adjusted for age and sex. Following additional adjustment for the Healthy Diet Indicator, this did not attenuate the inter-cohort difference at all (HR = 2.89 [2.25-3.71]). This suggests that diet was unable to explain international differences in outcome. This finding does come with some caveats. It may still be plausible for international differences in non-responders to be caused by variables such as diet. However, this still leaves unanswered the cause of variation among responders, and it is more plausible for the cause of international difference between responders and non-responders to be the same. Furthermore, the hypothesis of diet driving international differences is less compatible with the rapid changes seen in mortality just 1-2 years after the breakup of the Soviet Union. One would anticipate much longer lag periods between exposure to diet and changes to mortality.

1.3.2.3. Alcohol

Another early hypothesis for the cause of international divergence in mortality from 1991 onwards was alcohol.(162) Ecological data showed how short-term fluctuations in national rates of consumption correlated with similar fluctuations to mortality from cardiovascular diseases(163, 164) and all-cause mortality.(165) Rehm *et al.* again modelled alongside the Global Burden of Disease framework, to take exposure-disease associations from the West, and coupled this with estimates of higher prevalence of exposure and outcome in Eastern Europe, to calculate that male alcohol-related mortality is higher in Eastern Europe when compared to Western Europe.(166) However, the absence of individual level data in

such ecological and modelling approaches leaves open the possibility that their findings are confounded by third factors, which cause both higher rates of disease and alcohol. For example, if psychological stress is the underlying cause, then removing alcohol may not prevent CVD. Using individual-level data, case-control and cohort studies from Russia have associated some subtypes of alcohol with CVD outcomes. However, each of these studies did employ somewhat unconventional analytical techniques, as discussed below.

The first case-control study by Leon *et al.* initially recorded alcohol consumption from a range of sources (incl. beer, wine, spirit, and non-beverage sources, such as specialty perfumes).(167) The existing literature might suggest that one solution is to consider the primary exposure as *Total alcohol consumption*, from all these sources combined. However, as associations for this outcome were not reported, it is possible that such associations were not distinguishable from occurring by chance alone. Instead, the paper reported large mortality associations (unlikely to happen by chance alone) exclusively for three exposures: non-beverage alcohols; one type of spirit (vodka); and social features of problematic alcohol dependence (defined as having one or more episodes of *zapoi* in the past year [a cultural phenomenon more common in Eastern Europe, where a person is continuously drunk for >2 days] and/or at least twice weekly occurrence of excessive drunkenness, hangover, or going to sleep at night clothed because of being drunk). Altogether, the prevalence of all of these risks might be approximately 15-20% in the Russian general population. Coupled with a large Odds Ratios for death in men aged 25-54, this analysis suggested that as much as 43% of premature male mortality in this age group might be attributable to these subtypes of alcohol. However, the presumed absence of effect from total alcohol consumption makes possible that these results were confounded by residual socioeconomic status (which was statistically adjusted relatively simply, as just three categories of education), smoking (which was statistically adjusted as five categories), or other psychosocial variables (such as material deprivation, unemployment, depression and/or social

isolation, neither of which were measured). Data from other high- and low-income countries suggests that higher SES individuals are more likely to consume more alcohol.(168) At first, this observation may seem to be against my critique of the Leon *et al.* study (since SES is not a plausible confounder of the association between *total alcohol consumption and mortality*). However, the SES gradient appears to reverse when one considers other measures of alcohol behaviour, such as binge drinking. This now makes it plausible for SES to be a confounder of the association between *problematic alcohol behaviours* and *mortality*.

Secondly, these exposures were collected from proxies after death. It is possible that proxy informants up weighted recollections of alcohol and down weighed recollections of these measured confounders. For example, smoking behaviours may be less visible than alcohol behaviours, thereby inducing more measurement error, which leads to overinflating the attribution made to alcohol.

A subsequent case-control study by Zaridze et *al.* dealt with the problem of differential reporting bias among proxy informants, by setting the control group to not be those who are still alive, but those who died from cause plausibly unrelated to alcohol.(169) The investigators reported associations with *total alcohol consumption*, for a range of mortality outcomes. Critically, however, their analysis made no adjustment for education, marital status, ethnicity, job loss, or income, despite reporting in the same paper that these variables were collected at baseline. Furthermore, while smoking history was originally measured in much detail, this variable was dichotomized (yes/no) during the analysis stage, thereby substantially increasing residual confounding from smoking. In a prospective follow-up of the same baseline study, Zaridze *et al*. omitted associations with *total alcohol exposure* and presented instead exposure for *vodka consumption*.(170) In this paper, the statistical analyses made finer adjustments for smoking and education, but did not adjust for marital status, ethnicity, job loss or income.

To summarize, three articles in a reputable journal (*The Lancet*) make a convincing case that subtypes of alcohol are strongly associated with alcohol in a relatively crude analyses. However, only one paper out of three presents results for total alcohol consumption. This leaves open the possibility that it may not be exposure to ethanol molecules per se, but rather the socioeconomic confounders of certain types of alcohol, that causes both mortality and alcohol behaviour. The argument of a large causal effect from alcohol to mortality could be strengthened if those analyses could be repeated, where *total* alcohol consumption is considered as the primary exposure, and where the various putative confounders that have been measured around smoking, socioeconomic and psychosocial factors are rigorously adjusted for. This withstanding, if alcohol is a causal mediator of the association from SES to mortality, then policies that lower alcohol consumption should reduce health inequalities, which would be praiseworthy outcome. This is consistent with a report by Mackenbach *at al.*, where an increase in alcohol affordability over time was associate with an increase in educational inequalities in alcohol-related mortality, over time.(171) Nonetheless, the contribution made to inequalities in total mortality (at age 35-79) were a modest 2-9%, suggesting that the mediators of socioeconomic differences in mortality, as well as the cause of temporal and geographical variations, remain to be determined.

Prospective analyses from the HAPIEE cohort reported associations between total alcohol intake >60g/day and mortality from CHD, CVD and all-causes.(172) This level of drinking is equivalent to 3.8 times more than the UK Chief Medical Officer guidelines, making it quite a rare exposure. When consumption was categorized into various groups below this threshold, then these were not consistently associated with any mortality outcome. These associations for >60g/day in men were attenuated by 29%, 40% and 56% after adjustment for a range of behavioural, socioeconomic and psychosocial factors (incl. education, marital status, economic activity, material possessions, economic hardship, depressive symptoms, smoking, physical inactivity, BMI, and prevalent CVD and

cancer). There were no associations with other measures of alcohol, such as binge drinking, drinking frequency or drinking pattern. Since exposure to >60g/day of alcohol was so rare (2.8% prevalence), the population attributable risk fraction among men was also only 2.8%. Additional sensitivity analyses suggested that, even if the point estimate was biased by twofold, and even if non-responders had 13x greater prevalence of exposure, and even if non-responders had twice the hazard ratio of responders, then the population-attributable fraction for all-cause mortality would only come as high as 22% (what can only be considered an extremely upper margin of the likely underlying estimate). These calculations could be further enhanced, in future analyses, by also accounting for borderline significant associations among those who drink between 1 to 3.8 times above the recommended safe limit (for example, by using polynomial splines to model a nonlinear exposure-outcome associations). These refinements may indeed be warranted, in light of emerging evidence from Mendelian randomization studies, which suggest that alcohol may indeed be a causal risk factor in the development of CHD.(173) Nonetheless, it seems unlikely to me, that total alcohol consumption in the HAPIEE study would be associated with more than 50% of all-cause mortality in the adjusted analyses, even after additionally accounting for the hazard from lower levels of consumption.

1.3.2.4. Psychosocial and socioeconomic factors – area level data

It has been hypothesised that the transition from communism to capitalism in Eastern Europe could have exacerbated the influence of psychosocial hazards on CVD.(155) More specifically, a greater proportion of the total population may have become exposed (when compared to the proportion who are exposed in countries from Western- and Central-Europe). Alternatively, the magnitude of presumably causal effects could have increased in Eastern Europe. During the 1990s, countries of the former Soviet Union experienced a recession many times larger than the Great Recession of 2008. In Russia, the period 1990-1996 saw GDP per capita (constant, adjusted for purchasing power parity), fall by 40%. Even by 2015, it was only 17% higher than in 1990, in contrast to a 65% increase between 1990-2015 for the World average.(174)

When a recession of unprecedented magnitude is coupled with vanishing social welfare, rising crime, political uncertainty, and changing cultural expectations, it is only fair to call the speed and magnitude of net social change as literally otherwise unprecedented during modern times. One can speculate that this might have led to persistent and unresolved anxiety about the future, and possibly feelings of collective failure, shame and loss of purpose. This might result in greater exposure to low perceived control, greater depressive symptoms, greater material deprivation and perceptions of loss in social status. This hypothesis has received very little empirical study.

Within this region, countries whose residents report greater perceived control also have lower levels of mortality, in a cross-sectional analysis.(175) However, as this analysis did not adjust for putative confounders, a range of alternative causal chains could be constructed to account for this (e.g. alcohol/diet → high mortality → low perception of control). This period of massive social change coincided with large increases in suicide and CVD mortality. Formal time-series analyses, where suicides are taken as a proxy marker of stress and CVD is the outcome, have not been conducted.

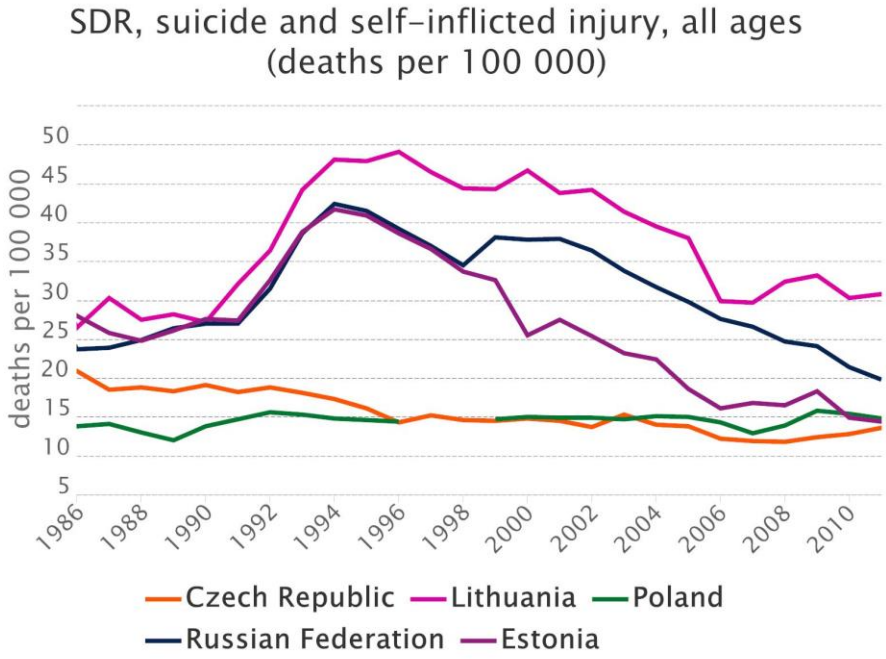


Figure 8. Temporal trends in suicide mortality, in selected countries from Central and Eastern Europe.
Source: https://gateway.euro.who.int/en/hfa-explorer/#9aZl40c8fU
Data: WHO Health For All Database

Temporal changes to income inequality have been correlated with temporal changes to life expectancy at the population level, when comparing countries ecologically.(175) However, this analysis did not adjust for putative confounders such as GDP per capita, or country-level fixed effects. Despite these weaknesses, this observation is consistent with other ecological and econometric analyses that have found an association between countries who chose to privatize their state-controlled industries quicker, and corresponding higher mortality.(176)

One age-period-cohort analysis suggested that it may not be purely *period* effects that are associated with mortality, as the literature has assumed so far, but in addition *cohort* effects may be seen for that cohort of men who lived through much of the hardship of the second World War and

subsequent Stalin's regime.(177) In other words, it was not the social change of the 1990s per se which caused an increase in mortality, but instead this happened to coincide with the increase in mortality that would have happened, regardless of whether the Soviet Union collapsed or not.

The *"Global Burden of Disease Study 2015"* study constructed a measure of Socio Demographic Index (SDI).(178) This ecological indicator is a composite of income per capita, educational attainment, and fertility, and is seen as proxy for general societal advancement. What I found striking in their analysis is how dramatically the life expectancy in Eastern Europe deviates away from an otherwise close correlation, when world regions are plotted so that each data point is a particular point in time. While this analysis captures temporal trajectories, it is most suited to thinking about over- and under-achievers in terms of mortality outcomes, given a fixed level of societal development.

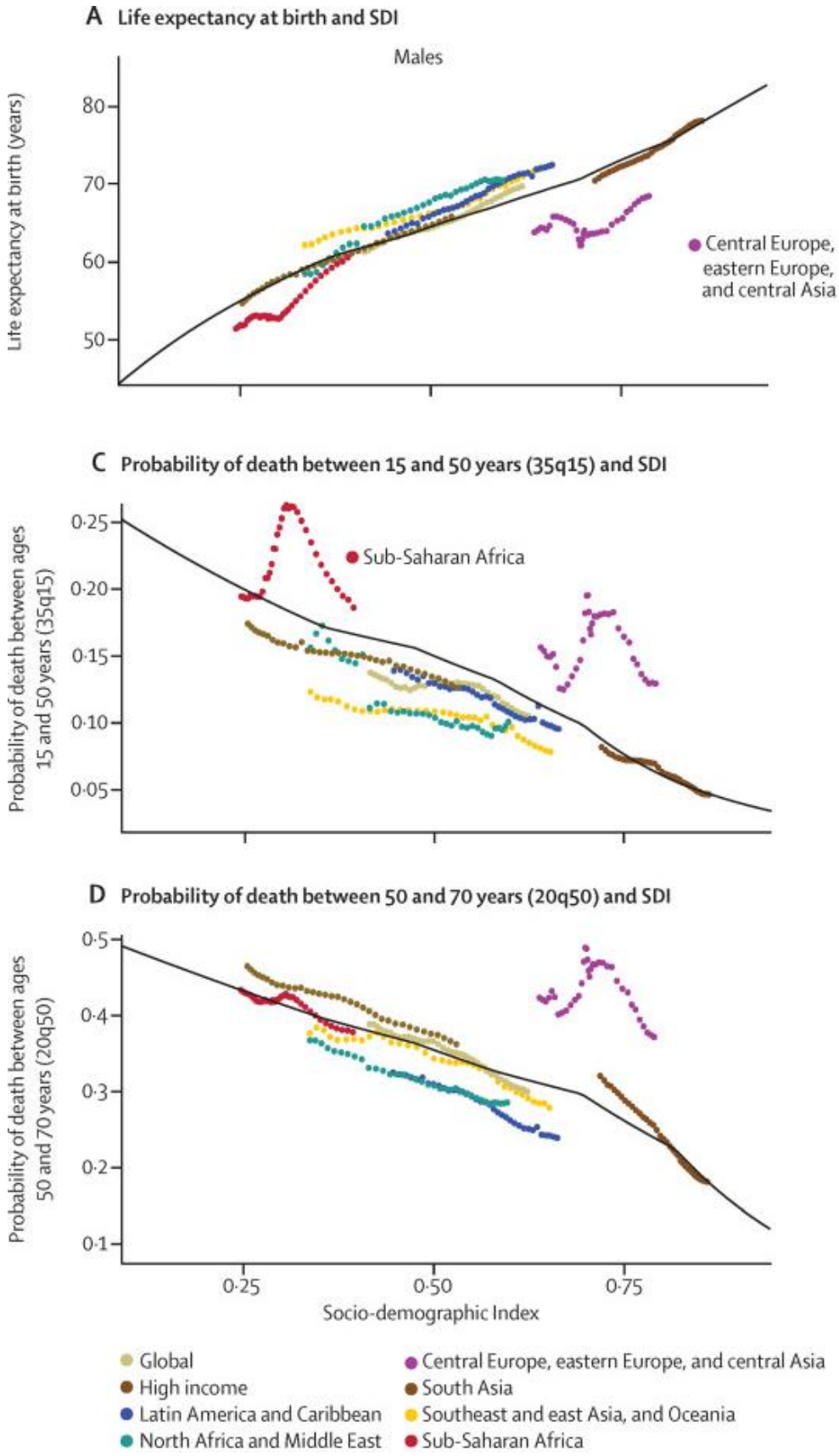


Figure 9. Association between Socioeconomic development and male mortality, among the Global Burden of Disease super-regions between 1980 to 2015. *Adapted from (178).*

The most revealing observation from figure 9 is that if countries in Eastern Europe increase economic productivity, educational attainment and lower fertility, then this may not be sufficient to close the international gap with other high-income countries. Personally, I dislike the coupling of these disparate measures of “development”, as political discussions rarely give as much focus to educational attainment or fertility, as they give to economic productivity. Moreover, Eastern Europe may be a deviant also in these facets: while levels of educational attainment are nominally quite high, the quality of educational qualifications may deserve additional attention. Similarly for fertility: low rates of fertility may be interpreted as a state of high-development (as suggested by the GBD authors). An alternative interpretation would be one of uncertainty and shock, since fertility rates tend to fall at times of recession or other times of large social change.(179) In this context, future research efforts could concentrate on the association between economic growth (and perhaps inequality therein) and mortality, while attempting to adjust for confounders like fertility and perhaps also education.

1.3.2.5. Psychosocial and socioeconomic factors – individual data

The area-level studies above are better able to identify factors that have a homogenous association with health, for all residents of that population. In other words, if income inequality causes mortality, and if the magnitude of effect is homogenous regardless of the whether the individual has low or high health behaviours, then area-level analyses would be able to detect this. However, this method has two weaknesses. First, area-level studies are less able to capture various confounders (such as a different pattern of drinking or eating in one country, but not the next), particularly as an analysis of Eastern European countries will usually include fewer than 20 countries, thereby limiting statistical power. In comparison, individual-level data may be better placed to control for measured confounders, thereby enabling analysts to produce a less biased estimate of whether their primary exposure may account for temporal or international variation.

Second, area-level studies cannot comment on how any temporal variation may be socially patterned.

My review of studies, that have investigated whether psychosocial or socioeconomic factors may account for temporal or international mortality differences in Eastern Europe, resulted in three studies. Each of these analysed routine mortality data, and were therefore unable to adjust for potential confounders such as smoking and blood pressure. These have found that the rise in mortality in Eastern Europe during the 1990s was largest among single middle-aged men of low education.(4, 70, 180) Apart from these, I am not aware of other individual-level studies that have investigated whether international or temporal variation in mortality in Eastern Europe might be associated with psychosocial or socioeconomic factors.

However, my own reinterpretation of the data presented by Vandenheede gave me some early clues to be more sceptical about this hypothesis.(2) I looked at mortality rates of the most privileged men across the various country cohorts (whereby privilege could be measured through education, relative deprivation or absolute deprivation). Privileged Russian men had about twice the baseline mortality than equally privileged men in the other three countries. This suggested to me, that a large part of the mortality gap between Russia and Central Europe is operating *outside* of socioeconomic differences, at all levels of society (at least among the responders of the HAPIEE study).

1.4. Predicting total cardiovascular risk in individuals

Before evaluating the literature on the performance of existing CVD risk prediction models, it may be helpful to summarize the optimum practice for developing and validating a clinical prediction model, as well as how to best measure performance and other measures of their utility. This is particularly timely, since the best practice recommendations around model development have improved considerably in the last 10 years. This reflects the expansion in literature of both methodological simulation studies, as well as real life evaluations of risk prediction models. Accordingly, many of the popular models that are now recommended for clinical use (e.g. SCORE, QRISK2) were developed with methods that are now seen (with hindsight) to have some weaknesses that could be avoided when developing new models. Thus my review of existing models (and their performance) needs to be situated in a context of quickly changing norms and standards. Annex 2 provides a summary of current methodological thinking around the best practice for developing and evaluating risk prediction models. Key elements relevant to my review of the literature are mentioned below. Key elements relevant to why I chose a particular method are additionally mentioned in my Methods section 2.3.

1.4.1. History of CVD prediction models

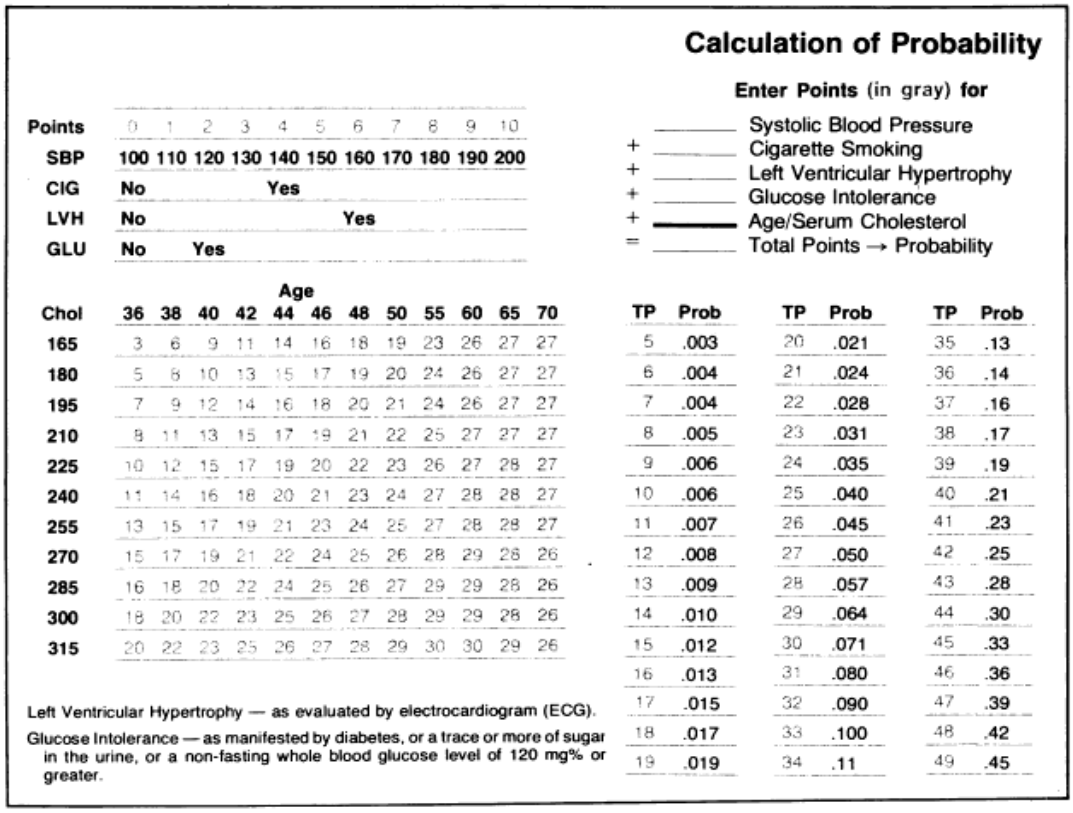
I will focus on models that have predicted the development of CVD (i.e. fatal/nonfatal CHD, +/- stroke, +/- other atherosclerotic CVD, +/- other non-atherosclerotic CVD), among participants free from CVD at baseline.

1.4.1.1. In Western Europe

The first CVD model in my opinion, was probably derived by Kannel *et al.* in the mid-1970s from the Framingham study (hereinafter the *"Fra-Kannel"* model).(181) This paper is remarkable for fulfilling many of the modern criteria for model development (incl. arguing against dichotomizing continuous variables; Fitting nonlinear effects such as age-squared in an appropriately theory-driven manner; Suggesting a suitable ratio of

predictors to events; and the accurate identification of 6 core predictors that later replicated in practically all subsequent CVD models).

In 1982 a refined version of the model was summarized onto a simple clinical reference card.(182) This innovation allowed continuous risk factors (such as age, blood pressure and cholesterol) to be categorized into around 10 categories each, so that clinicians could use them without a calculator. This card also was remarkable in operationalizing the age-cholesterol interactive term (figure 10), a feature that then disappeared from subsequent CVD models until its reintroduction in 2014. Curiously, this innovative publication received little clinical or academic attention (for example, by attracting just 19 citations over 35 years).



Calculation of Probability

Points	0	1	2	3	4	5	6	7	8	9	10
SBP	100	110	120	130	140	150	160	170	180	190	200
CIG	No				Yes						
LVH	No						Yes				
GLU	No		Yes								

Enter Points (in gray) **for**

_____ Systolic Blood Pressure
\+ _____ Cigarette Smoking
\+ _____ Left Ventricular Hypertrophy
\+ _____ Glucose Intolerance
\+ _____ Age/Serum Cholesterol
= _____ Total Points → Probability

Chol	Age 36	38	40	42	44	46	48	50	55	60	65	70
165	3	6	9	11	14	16	18	19	23	26	27	27
180	5	8	10	13	15	17	19	20	24	26	27	27
195	7	9	12	14	16	18	20	21	24	26	27	27
210	8	11	13	15	17	19	21	22	25	27	27	27
225	10	12	15	17	19	20	22	23	26	27	28	27
240	11	14	16	18	20	21	23	24	27	28	28	27
255	13	15	17	19	21	23	24	25	27	28	28	27
270	15	17	19	21	22	24	25	26	28	29	28	26
285	16	18	20	22	24	25	26	27	29	29	28	26
300	18	20	22	23	25	26	27	28	29	29	28	26
315	20	22	23	25	26	27	28	29	30	30	29	26

TP	Prob	TP	Prob	TP	Prob
5	.003	20	.021	35	.13
6	.004	21	.024	36	.14
7	.004	22	.028	37	.16
8	.005	23	.031	38	.17
9	.006	24	.035	39	.19
10	.006	25	.040	40	.21
11	.007	26	.045	41	.23
12	.008	27	.050	42	.25
13	.009	28	.057	43	.28
14	.010	29	.064	44	.30
15	.012	30	.071	45	.33
16	.013	31	.080	46	.36
17	.015	32	.090	47	.39
18	.017	33	.100	48	.42
19	.019	34	.11	49	.45

Left Ventricular Hypertrophy — as evaluated by electrocardiogram (ECG).

Glucose Intolerance — as manifested by diabetes, or a trace or more of sugar in the urine, or a non-fasting whole blood glucose level of 120 mg% or greater.

Figure 10. The first semi-A4 sized risk guide, to calculate CVD risk in clinical settings, from (182).

In 1991, Anderson *et al.* again slightly refined the Framingham models, and persuaded the American Heart Association to endorse them. This lead to their adoption by a wider pool of clinicians.(183, 184)

In 2003 The European Society of Cardiology published and endorsed the SCORE model.(185) This was the first attempt to offer one model for use across multiple countries. The European cohorts that were used to derive SCORE were separated into “low risk” and “high risk” regions, with each deriving its own model. It is unclear by what criterion they decided on the allocation of countries into these two categories. For example, figure 11 illustrates how there is a negligible difference in the incidence of the primary outcome between Sweden (a high-risk country) and Belgium (a low-risk country). Perhaps the authors were intending a 50:50 split in the sample size, as this turned out to be 56:44 for the low- and high-risk models, respectively. (This intention could be further criticized, since the high-risk model, having more events, was better powered for model derivation than the low risk model.) In essence, already from inception, the allocation of countries into “high” and “low” risk status appears relatively arbitrary, thereby limiting its real life calibration. This is further aggravated by the passing of time. Most of the CVD events occurred in the late 1980s, by which point international differences between Russia and capitalist European countries were not half as large as they were in the 2010s.

	Country	Cumulative CVD death rate by age 65*
Men	Finland	12.80%
	Russia	11.91%
	Norway	7.91%
	UK (BRHS)	7.11%
	UK (Scotland)	6.49%
	Denmark	6.44%
	Sweden	4.80%
	Belgium	4.79%
	Germany	4.72%
	Italy	4.01%
	France	3.20%
	Spain	2.81%

Figure 11. Derivation of the male SCORE model. Countries that contributed to the "high risk equation" are highlighted in red box, while countries that contributed to the "low risk equation" are not highlighted. *Adapted from (185).*

2007 saw the publication of the ASSIGN model in Scotland, which additionally incorporated family history and area-level deprivation.(186) Shortly thereafter, QRISK was published for England being the first and only CVD model derived from electronic healthcare records, during the period 2007-2017.(187) One of the strengths of the QRISK project is how models are derived in settings virtually identical to where they will ultimately be used, which is likely to lead to much better calibration when compared to deriving models from cohorts studies with greater selection bias. However, the QRISK project is only able to explore predictors that GPs already consider important and record in their notes. As such, it is not able to test the potential benefits from asking GPs to consider new predictors. In contrast to models from USA where the hazard of age increases with age (using an age^2 term) the English model included the opposite interaction (log-age). Perhaps this alteration could reflect how the

hazard of old age may additionally be captured by comorbidity markers, like Diabetes and Rheumatoid Arthritis, in QRISK.

One year later, QRISK was updated into QRISK2.(188) As these researchers had access to a large database (with 96 709 events) they added details of further comorbidities, as well as negative interactions for most risk factors with age. The latter could be a marker of latent gene-environment interactions, whereby smokers who survive until age 70 without CVD are unlikely to contract CVD due to smoking alone, after the age of 70. However, as cholesterol was missing for most participants, this prompted considerable methodological critique. In particular, when imputing cholesterol and omitting the outcome from the predictor matrix (as was originally done), the imputed cholesterol values were substantially different as opposed to using a predictor matrix that included the outcome (as was later recommended). From this high-profile example, the former is now recommended as standard established practice for dealing with missing data in prediction settings. Omitting the outcome biases the imputed cholesterols towards the mean, thereby inducing a type of regression dilution bias that attenuates the beta coefficient of cholesterol artificially downwards towards the null.

In 2014 the American Heart Association switched its endorsement away from the Framingham models, towards the Pooled Cohorts Equation (PCA).(189) This added to the Original Framingham and Framingham Offspring Studies three further cohorts: ARIC (Atherosclerosis Risk in Communities); Cardiovascular Health Study; and CARDIA (Coronary Artery Risk Development in Young Adults). Most of the participants were recruited in the 1980s, with events occurring during the 1990s.

A summary of the predictors used by these better established models is shown in table 1.

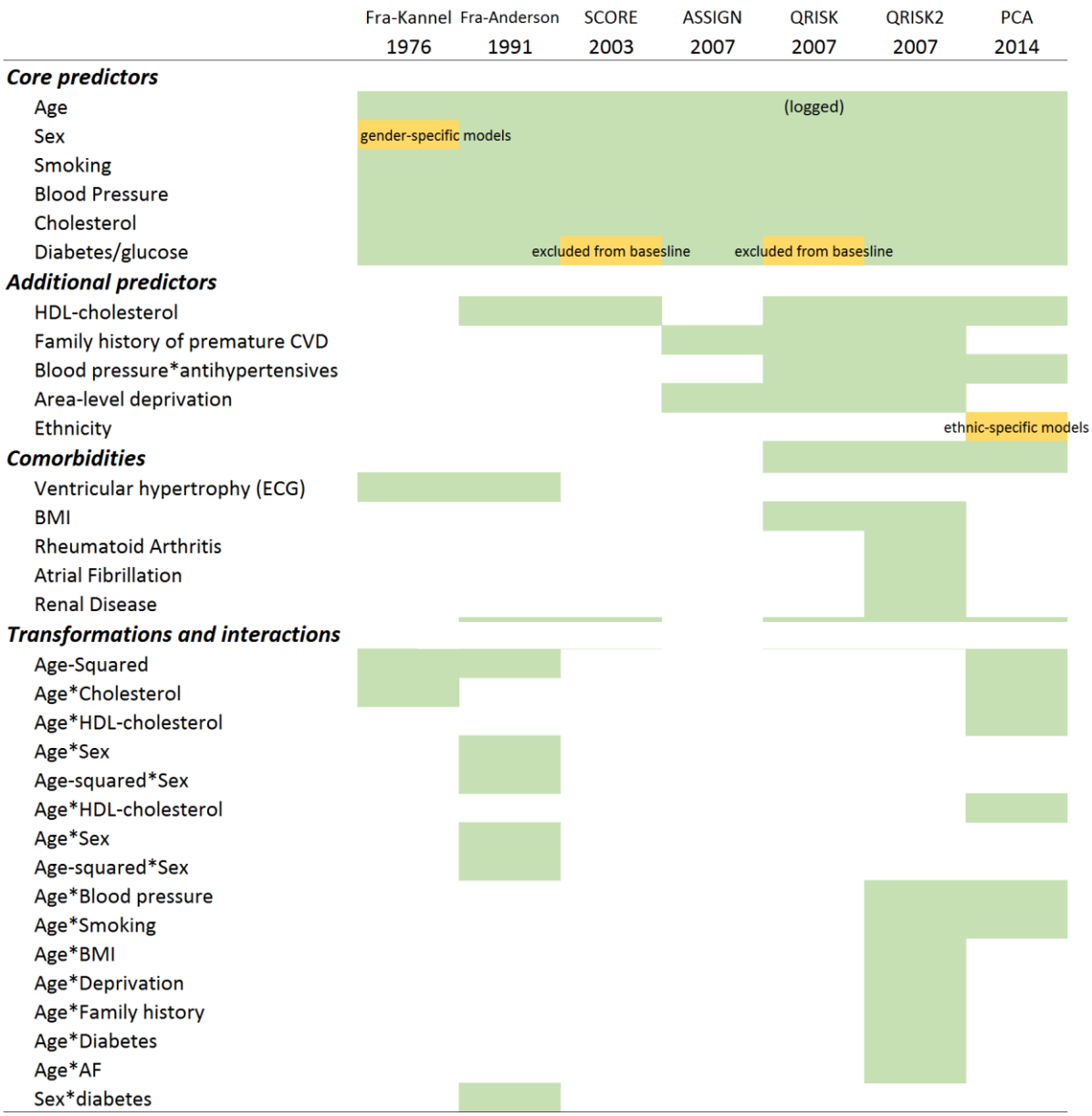


Table 1. List of predictors used in popular CVD risk prediction models from Europe and the USA
(Fra=Framingham; PCE=Pooled Cohorts Equation; ECG=Electrocardiography; AF=Atrial Fibrillation; BMI=Body Mass Index; HDL=High-Density Lipoprotein; CVD=Cardiovascular Disease)

Importantly, in each of these popular risk prediction models above, the publication which detailed the model derivation did not include any data about external validation. Best practice in model development suggests that external validation is probably one of the most rigorous steps one can take, to demonstrate that the models are not overfitted and will perform

similarly well when adopted to real-life clinical situations (Annex 2). The importance of this cannot be overstated, since most of the caveats and considerations around deriving accurate risk prediction models are centred around trying to minimize overfit (e.g. how to select variables, any departures from linearity, interactions, and ensure sufficient power). However, external validation is difficult to conduct, which is why it is rarely reported papers that derive new models. One external validation study found that the Framingham performed well in the UK among those of mid-to-high risk, but underestimated risk among those of lowest risk.(190) This is a relatively rare finding. In contrast, most external validation studies find how risk is overpredicted in the validation dataset, where events are less common than expected. This has happened in the USA after testing the PCA model.(191-193) Typically this is because the validation dataset is healthier, for example due to period effects (194), or alternatively due to overfitting.

Of note, early models like Framingham omitted socioeconomic status. If this model was implemented, then this would have widened socioeconomic inequalities in CVD.(195, 196) By same token, it is plausible that widespread use of an international model such as SCORE, which does not sufficiently capture the underlying difference in baseline rates between Russia and Western Europe, may widen international health inequalities, by encouraging underuse of preventative interventions in those countries of greatest risk.

1.4.2. Previous multi-country scores

As described above SCORE was the first attempt to provide risk estimates for multiple countries. However, it is applicable only to Europe, period effects have substantially weakened the models' utility, and they probably did not account sufficiently for heterogeneity in CVD rates across Europe.

A second attempt was made by the WHO in its guidelines on the assessment and management of CVD risk.(197) This was the first time that a global approach was taken, to provide an approximate risk model

for countries with limited prospect of deriving their own models. The project can be further congratulated on a further conceptual innovation: instead of taking the risk factor Hazard Ratios from the same dataset where the estimate of baseline risk is taken from, it was perhaps the first CVD model to draw these inferences from separate sources.

However, the application of this idea had substantial limitations. The baseline disease risk was taken from an early iteration of the Global Burden of Disease Study that had not been peer reviewed. The underlying data quality of this has been subject to much criticism. The WHO risk developers compounded this limitation by opting to predict not just fatal CVD, but also nonfatal CVD whose international data quality is extremely heterogeneous. Secondly, the risk factor Hazard Ratios (HR) were taken not from a single joint regression that involved all risk factors, but apparently some HRs were estimated from univariate models that had only been adjusted for age and sex.(198) This can lead to large errors in overestimating risk for persons with more than one risk factor, and subsequent problems with discrimination and calibration. As such, the WHO model has never been published in a peer-reviewed form and will be discarded in the rest of this thesis.

A third attempt was made in 2015 with the derivation of GLOBORISK, initially for 11 countries (198) which was later extended to 182 countries(199). Their second publication was published in March 2017, by which point I had completed most of the analyses later presented in this thesis. Therefore, this publication could be viewed as a parallel development to my own work.

The GLOBORISK project should be congratulated for continuing the vision of the earlier WHO model, but managing to translate this into a reliable model that could contend as the default option for most countries of the world. However, the GLOBORISK model contains two considerable limitations, the first around calibration and the second around discrimination.

In contrast to SCORE, GLOBORISK added nonfatal events to its outcome. This can be seen as a strength in terms of trying to bring the rest of the world to a state comparable to Western Europe and the North America, in terms of studies of interventions, guidelines and national programmes. However, adding nonfatal events also weakens the model, by introducing substantial inter-country heterogeneity in data quality. This may be the reason why the GLOBORISK website (http://www.globorisk.org) predicts higher risk of fatal/nonfatal CVD for male participants in Finland, when compared to risk estimates for equivalent participants from Estonia, Poland and the Czech Republic. This is inconsistent with the epidemiology of this region, as demonstrated in figure 12.

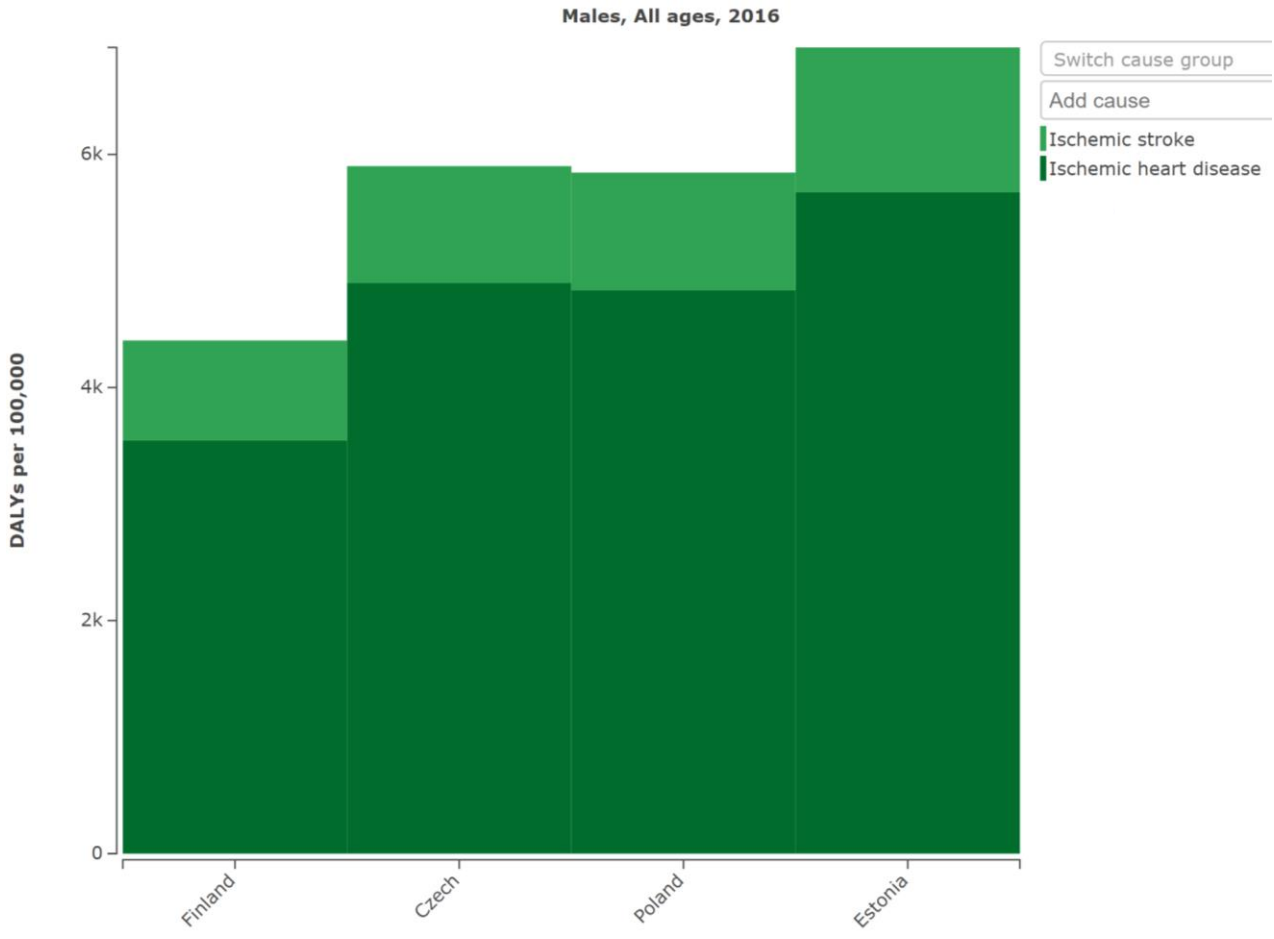


Figure 12. Comparison of the burden of ischaemic cardiovascular disease across four European countries.
The data describe the Disability-adjusted Life Year (DALY) burden for men of all ages in 2016. Data are from (200). This geographical pattern is not capitulated in the GLOBORISK CVD risk prediction model.

A second explanation for these implausible predictions is how the GLOBORISK predicts risk for the year 2024, using linear extrapolation from trends between 2001-2015. Trends in Eastern Europe have been chaotic, and there is little evidence to suggest that these extrapolations will hold true. Finally, although the GLOBORISK project set out with a focus on improving calibration, no calibration data are presented in either paper nor its appendices, other than the largely discredited Hosmer-Lemeshow χ^2 test. To summarize this first limitation, I would have found the GLOBORISK project much more useful, had it focused exclusively on predicting fatal CVD events, and had it demonstrated calibration.

As a second limitation, the authors assumed that the HRs found in the USA are generalizable to the other 181 countries. There is very little prospective cohort data to empirically test this claim. The authors cited one analysis, where HRs from Australia/New Zealand were compared against “Asia”. Small differences were found for some HRs (such as blood pressure and cholesterol).(201) However, this study amalgamated data from primarily China and Japan into one large region, despite the fact that these two countries have very different profiles of economic development and CVD. The GLOBORISK team secondly cite the INTERHEART study as evidence for the homogeneity of Odds Ratios between regions. This study explored the range of Odds Ratios for 6 risk factors, across 11 regions (i.e. 60 comparison groups). Across all 6 risk factors, there appeared to be considerable evidence that heterogeneous Odds Ratios were unlikely to arise from chance alone (figure 13).

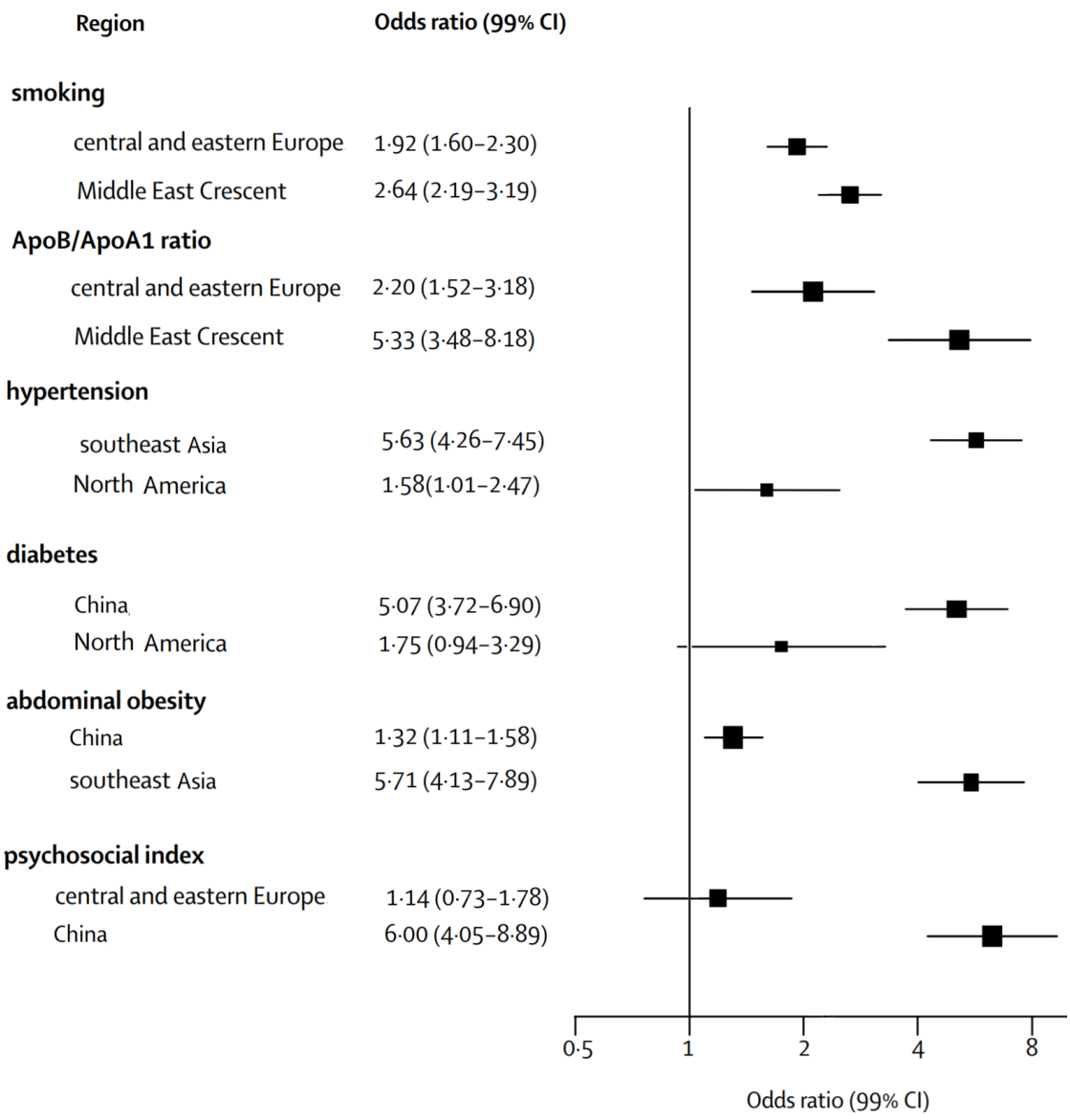


Figure 13. Regional variation in Odds Ratios for six risk factors for acute myocardial infarction, from the INTERHEART study. *Confidence intervals are 99%, while 95% confidence intervals will be narrower. Adapted from (202), where I have purposefully selected those regions with the greatest heterogeneity.*

However, the INTERHEART authors did not explore this observation statistically, for example no significance tests nor Bonferroni corrections are reported. It is also possible that the decision to present 99% Confidence Intervals, as opposed to the established practice of 95% Confidence Intervals, may have been with the intent of projecting an illusion of homogeneity to those who were unaware of this detail.

Therefore to summarize, at current there are no reliable analyses from which to infer whether it is appropriate or not to extend Hazard Ratios and other risk factor coefficients from Western European/North American studies onto other regions of the world. The answer to this question informs the design of risk prediction models elsewhere, but also estimates of global burden of disease, and priority setting for national public health authorities.

1.4.3. CVD prediction in Eastern Europe

Vikhireva *et al.* dichotomized the high-risk SCORE at threshold of >5% predicted risk of fatal CVD, and applied this to 8 MONICA and HAPIEE cohorts from Eastern European countries. Calibration was good for the earlier MONICA Central European samples, with Predicted/Observed (P/O) ratios ranging from 0.9 to 1.3. However, calibration became entirely useless in the more recent HAPIEE sample (with P/O ratios ranging from 3.0 to 7.8).(18) In other words, Applying SCORE predictions to men from the Polish HAPIEE cohort would lead to 8 times as many predictions, while in reality there was only one event. This has the potential to lead to overtreatment, side effects, and resource waste.

This mismatch between MONICA and HAPIEE is probably driven by three complimentary mechanisms. First, period effects in Central Europe led to lower baseline rates of disease by the time of the HAPIEE study, thereby lowering observed rates. However, this alone may not account for the eight-fold difference in calibration. Second, it is likely that the MONICA sampling frame was more similar to the sampling frame used in the other cohorts used to derive SCORE, while HAPIEE may have oversampled higher SES participants comparatively more, with lower baseline rates of disease. Third, there is a secular trend across the world, whereby response rates were highest during the 1970 (i.e. the derivation of SCORE), slightly lower in the 1980s (i.e. MONICA) and lower still in the 2000s (i.e. HAPIEE). It is plausible that non-response bias, which tends to produce artificially low mortality rates, is greater in more recent time periods.

Vikhireva *et al.* also reported extremely low discriminatory power of the SCORE model across all cohorts, with C-statistics ranging from 0.54 (very close to a random coin flip) to 0.74. There was some indication of improvement in discrimination over time (the average C-statistic in MONICA was 0.63, Vs. 0.67 in HAPIEE). Even after restricting analyses to HAPIEE data only, the C-statistic was lower among men than women (0.63 Vs. 0.70), while men face much higher needs for accurate CVD prediction.

Jdanov *et al.* published a SCORE model that was recalibrated to the Moscow and St. Petersburg populations. This is unlikely to generalize to the wider Russian population, on account of higher SES and better health outcomes in larger cities. Its use elsewhere would lead to undertreatment. Their model was based on 652 events detected between 1975-2009.(203) Unfortunately, the publication did not present measures of calibration, discrimination or reclassification. I am not aware of other English-language publications that evaluate the performance of CVD risk prediction tools in Eastern Europe.

1.4.4. Attempts to augment models with new predictors

Attempts to date to improve CVD risk prediction with new risk factors have on the whole been disappointing. Usually, these efforts are focused around improving discrimination, with some papers additionally reporting reclassification.

Biomarkers such as CRP, troponin, and fibrinogen have typically improved C-statistics by around 0.003 to 0.010.(204) Downstream markers that are more specific to latent cardiac pathology may to have greater potential to improve discrimination. For example, BNP improved C-statistic by an outstanding degree - 0.08 units in one study(204), but only 0.011 in another(205). Coronary Artery Calcium Score (CAC) improved C-statistic by a similarly outstanding 0.05 units.(206) This study also improved Net Reclassification Improvement (NRI, discussed methodologically in Annex 2, section A2.2.10) reported across three categories by 0.25, which was

driven primarily by improvements among cases. However, this study had 15 events per parameter, shrinkage was not performed, and data were not externally validated. This makes it possible for chance overfit to account partly for these findings. The same authors also studied improvements from CAC in the same cohort again, this time focusing on improvements exclusively among those of intermediate Framingham risk (5% < FRS < 20%).(207) This improved C-statistic by 0.124, as well as a categorical formulation of Net Reclassification. However, the baseline risk model had an unusually poor fit (C-statistic = 0.642). Furthermore, with an event-to-parameter ratio of 8, no shrinkage of coefficients, and no external validation, a part of this is may again reflect statistical overfit.

1.4.4.1 How good is good enough?

Many of the studies in this field are difficult to compare with one another. Much like elsewhere in epidemiology, there may be slight differences in the population, response rate, definitions to risk factors or outcomes, or follow-up. However, studies of improvements to risk prediction are made less comparable by the use of a range of various reclassification measures. For example, one study may report reclassification across the entire risk gradient, a second study may dichotomize risk at a particular threshold, a third study may dichotomize risk at another threshold, a fourth study may use multiple risk categories, while a fifth study may look at multiple risk categories after having restricted the sample to those of intermediate risk. Changes in C-statistic are less prone to such inter-investigator variation, but are nonetheless sensitive to the ability of the baseline model to discriminate.

In the absence of guidance, I will consider the following interpretations throughout this manuscript. I will assume that baseline conventional CVD models all discriminate to similar levels (i.e. C-statistic of between 0.70 to 0.80). The following guidelines are informed by the observation that omitting blood pressure from a conventional risk prediction has resulted in a ΔC of only 0.004.(110)

ΔC < 0.004 = no improvement

ΔC ≥ 0.004 = minor improvement

ΔC ≥ 0.01 = substantial improvement

ΔC ≥ 0.02 = impressive improvement

ΔC ≥ 0.03 = outstanding improvement

1.4.4.2. Genetic predictors

Abraham *et al.* evaluated the utility of adding a genomic-risk-score, incorporating beta coefficients from 49 310 SNPs, to two conventional CVD models across three external cohorts. This improved C-statistics substantially - by 0.011 to 0.017 across the 6 comparisons. NRI, measured across 4 categories, was between 3 to 15% across the 6 comparisons, with continuous NRI being between 25 to 37%.(208) In contrast to this, previous studies that have included less than 50 SNPs which all met conventional levels of GWAS-significance have been much more disappointing (ΔC = 0.003, continuous NRI = 17%).(209, 210) This is consistent with current advice about selecting predictors into risk prediction models, which advise against using conventional thresholds (like $p<0.05$) to evaluate if a predictor should be included, but advocate for more relaxed thresholds.

To my mind, one large weakness in the current paradigm of genome-wide association studies is how they tend to adjust for just age and sex, and do not routinely adjust for conventional risk factors. This means that although readers and users of the top hits tend to think of these as “genetic signals” which can be compared and contrasted against “conventional risk factor signals”, they may in fact me measuring the same thing. That is, a seemingly cardiac SNP may actually be a mere marker of cholesterol, blood pressure, smoking and/or glycaemia. Accordingly, some high-profile applications of such “top cardiac SNP” have been interpreted to mean that genetics adds incremental value over and above conventional risk factors. However, some of these papers have not even defined what they mean by *“conventional risk factors”*, but merely claim to adjust for these in a multivariable manner.(211) The field could be improved by a rigorous

examination of cardiac GWAS signals, before and after adjustment for conventional CVD risk factors. If the cardiac GWAS do not attenuate, this would suggest that they are indeed independent. If they do attenuate, this would suggest that minor gains to prediction could be made either by measuring or modelling the conventional risk factors more precisely (which may turn out to be more cost-effective), or by using genetic proxies.

1.4.4.3. Physical inactivity

The benefit of adding physical inactivity to prediction models has been substantial in two studies (one among healthy participants, the other among participants with prevalent disease)(212, 213), and null once.(214) One study added physical activity, its interaction with gender, waist circumference, and peripheral artery disease, to find this package substantially improved the C-statistic.(215) However in another study, the addition of added physical activity, diet, alcohol, BMI and waist circumference made no improvement to C-statistic.(216) I am not aware of any study that has derived a model that includes physical inactivity, which has subsequently been externally validated.

1.4.4.4. Psychosocial and socioeconomic predictors

There are two motivations for including socioeconomic risk factors to risk prediction models. First, it may improve reclassification in the entire cohort, thereby fostering personalized preventative medicine for the population as a whole. In case of equitable access to, and uptake of healthcare, this has the potential to reduce healthcare inequalities. Second, even if socioeconomic risk factors do not improve reclassification in the entire cohort as a whole, they may still reduce health inequalities. Consider a new model that adds SES, whereby all those of low SES are given a very large risk coefficient. This could improve True Positive rates, sensitivity and reclassification among cases, but this could be offset by worse False Positive rates, specificity and reclassification among controls. There may be no change to unadjusted metrics, such as the C-statistic, or NRI. However, more statins may nonetheless be redirected towards those

with lower SES. In order to accurately predict whether the cost-to-benefit of such a divergence could be accompanied by a change in health inequalities (after considering the benefits and side effects of statins), these costs and benefits would need to be appropriately weighted (such as by using Net Benefit and Decision Curve Analyses). My thesis nonetheless, will ignore the aim of measuring health inequality, and will instead focus on the prospect of improving average health.

Table 2 presents an overview of 10 studies that have added a socioeconomic/ psychosocial risk factor, and reported either ΔC or NRI. (12, 18, 110, 212, 214, 217-221) Other studies have also reported the effects of adding SES variables, but these studies did not report either C-statistics, NRI, Net Benefit or Decision Curves.(222) All of the studies to date have looked at changes taking place within one cohort, and few have replicated their findings in external cohorts. Another common limitation is how most measures of performance have not considered clinically relevant improvements across clinically relevant risk thresholds (e.g. Binary NRI, detailed in Annex 2), but instead present measures that are sensitive to improvements at any risk threshold (e.g. Continuous NRI, C-statistic).

	Fiscella	Ramsay	Ingle	Vikhireva		Pujades	Ferrario	Veronesi		Schnohr	Graversen	Colantonio	
	2009	2011	2013	2014		2014	2014	2014		2015	2016	2017	
Description of the study													
Reference model	Framingham risk score	Recalibrated Framingham (& SCORE)	Recalibrated Framingham	SCORE		Age-sex		Recalibrated SCORE				Recalibrated PCE	
Added predictors	Education & Low income	Social Class	Psych. distress (GHQ-12)	Education	Marital Status	Area deprivation	Education		Family history	Vital exhaustion	Education		Income & Marital Status
Predictor form	Binary	Unclear	Unclear	Binary		4 quartiles	3 categories	Binary		4 categories	5 categories	Binary	Binary
Outcome	CHD death		CVD death	ASCVD death		Fatal/ nonfatal MI	Fatal/nonfatal MI/stroke			Fatal/ nonfatal MI	ASCVD death	Fatal/ nonfatal MI/ stroke	
Is risk categorized (for C-statistic)	No	Yes, three levels (6%; 20%)	No	Yes, binary (at 5%)		No	No	Yes, three levels (10%; 20%)	No	No	No	No	No
Results from the study													
ΔC-statistic / AUC	0.04	not reported	0.011	0.02	0.02	0.007	0.004	0.004	0.007	0.004	0.01	0.004	<0.001
Continuous NRI							0.15			0.18	0.03	0.16	<0.07
Categorical NRI		0.0018						0.03	0.03				
(among intermed.)							0.06	0.064	0.157				
Other results	Low-SES people would receive more statins		Physical inactivity ΔC=0.014			No interaction between SES- and- traditional risk factors	Greater benefits in Eastern Europe			Blood pressure ΔC=0.004; NRI=0.18	Vital exhaustion ΔC=0.003	No additional benefit from area-level deprivation	
Quality of the study													
Multiple imputation for missing data?	No	No	No	No	No	Yes	No	No	No	No	No	Reported in supplement	
Events per parameter?	3	59 to 92	17	9	795	43 to 162	36	36	192	32	27	25	25
Derivation & Validation in different	No	No	No	No	No	No	No	No	No	No	No	No	No

Table 2. Summary of methods and findings from 10 studies, that evaluated the addition of psychosocial/socioeconomic risk factors to CVD risk prediction models. *(Cont.=continuous; Cat.=Categorical; Intermed.=Intermediate; PCE=Pooled Cohorts Equation; NRI=Net Reclassification Improvement)*

Broadly speaking, of these 10 papers, the methodologically more robust studies have found no or little improvement, while methodologically weaker studies have found good improvement. This might suggest the potential for publication bias. Furthermore, I am not aware of a single study that has found improvement, and thereafter tested this beta coefficient in an external dataset. Another weakness of these studies is that they have typically dichotomized their exposure thereby discarding available information, although more granular categories may be easy to model.

Only one paper out of these 10 reported the benefit of adding up to three risk factors simultaneously.(221) (One Danish study did evaluate 15 novel predictors, but it only reported improvements in univariate models(214).) By analogy to the work with genetic and genomic markers, it may be more fruitful to add lots of markers simultaneously. Vikhireva *et al.* evaluated the benefit of adding education and marital status to the SCORE model.(18) They fitted 16 separate models stratified by each country, cohort and gender. This means that all the female models were substantially underpowered (with an event to parameter ratio ranging from 2 to 12, mode=4). In conjunction with no shrinkage and no external validation, this increases the risk of overfit substantially, so I will instead concentrate on their 8 male models (where the event to parameter ratio ranged from 4 to 25, mode=9), and report the modal statistic for each analysis. They found impressive improvements to discrimination following the addition of education (modal $\Delta C = 0.02$), marital status (modal $\Delta C = 0.02$), and that these improvements seemed to not to overlap with one another after adding both factors simultaneously (modal $\Delta C = 0.04$). With a similar methodological approach, Veronesi *et al.* evaluated the benefit of adding education and family history,(220) and also found these two components to contribute independently when added together.

In contrast to measures of SES, which has been the main focus of these 10 studies, there has been comparatively little focus on the contribution of psychosocial factors. Only one study has reported on the impressive

improvement after adding marital status. I am not aware of another study that has evaluated the addition of employment status or depression into an established CVD risk score. However, the addition of vital exhaustion has led to no improvement,(214) or minor improvement(110) in the C-statistic. Another study found substantial improvements after adding psychological distress.(212)

1.4.5. Summary of risk prediction

A decade ago, there was little need for clinicians to predict total risk, as each risk factor was thought to have a complimentary, reductionist effect. Recent thinking has shifted towards taking a more holistic approach. For example, statins appear to create additional off-target benefits, which do not involve cholesterol. This has created a growing clinical appetite for better risk prediction models. The methodology of how to create these is developing quickly. Basic models to predict cardiovascular disease have been around for over 40 years. However, there are considerable questions about how valid these models are to real-life populations from where their cohorts are drawn (due to selection bias and period effects); and even bigger challenges with exporting such models to other countries that lack their own derivation data. Accordingly, one avenue for development is the refinement of CVD models for countries where they are currently lacking. Eastern Europe has the world's highest burden of CVD, but both the SCORE and GLOBORISK models available there contain considerable weakness in calibration and discrimination.

A second avenue of development is the addition of psychosocial and socioeconomic risk factors.(223) A handful of studies that have tried this, with some finding impressive improvements while other not. None of these studies have externally validated their findings in another cohort. No study has evaluated the benefit of depression or employment status. No study has evaluated the benefit of adding four psychosocial/socioeconomic factors at the same time. And finally, no study has summarized *clinically relevant* benefits and harms in a single metric, such as Net Benefit.

1.5. Summary of the literature review

A large body of observational epidemiological studies, predominantly conducted in countries of Western Europe and Northern America over the past 50 years, have described a replicable and sizable associations between educational status, material deprivation, unemployment, social support, personality and depression in their ability to predict the incidence of cardiovascular disease among healthy middle-aged participants. At current, it remains relatively unknown to what extent they may mediate each others' effects, as well as to what extent similar associations may be found in Eastern European countries.

Data from twin/sibling studies, as well as natural experiments, are on balance more compatible with an interpretation of causal effects from education to all-cause mortality. Despite this, there remains considerable scepticism about whether changes to socioeconomic risk factors can cause changes to the incidence of cardiovascular disease.

Eastern European regions, and particularly ex-USSR countries like Russia, have substantially higher rates of cardiovascular disease when compared to countries in Central and Western Europe. The causes of this remain relatively unknown. The IMPACT model and MONICA studies suggest that conventional risk factors and healthcare may have played a smaller role in accounting for temporal changes in CVD. However, these studies have not investigated the causes of divergence among the ex-USSR countries after the breakup of the USSR. Dietary factors and alcohol have been proposed, but the data so far does not suggest that these can account for international differences observed in the 2000s. Area-level analyses have implicated rapid privatization, income inequality and low perceived control with higher mortality rates. However, these analyses cannot use individual level data to adjust for confounders. Existing individual-level analyses have implicated single marital status and low education as two factors that correlated with earlier death, however these analyses too did not adjust for confounders.

Models to predict cardiovascular disease have been around for over 40 years. However, these may not be accurate in Eastern Europe. Secondly, few studies have investigated whether psychosocial and socioeconomic risk factors can improve risk prediction. Where this has been reported, there has been no quantification of whether these benefits are clinically relevant, and there have been no subsequent reports of external validation.

In terms of translational impact, there is currently quite limited coverage of psychosocial and socioeconomic risk factors in medical textbooks and clinical guidelines. Their assessment is beginning to become a routine part of some aspects of public health practice, however this is not the case among primary care physicians and cardiologists. The causes behind this are numerous. There is potential for this to change over time, which in turn could be assisted by:

1) Clarifying the causal and mechanistic nature of the associations between psychosocial/socioeconomic factors and cardiovascular disease.

2) Using psychosocial/socioeconomic factors to explain why the burden of cardiovascular disease is inexplicably high in some countries of the world.

3) Demonstrating the utility of psychosocial/socioeconomic factors in improving the prediction of cardiovascular risk in individuals.

1.6. Aims and objectives

1.6.1. Overall aims

1. To investigate whether education is a **causal** risk factor in the aetiological development of coronary heart disease.
2. To investigate what could **mediate** the presumed causal association between psychosocial/socioeconomic factors and CVD.
3. To investigate whether the higher burden of CVD mortality found in **Russia**, when compared to Central European countries, could be attributed to psychosocial and socioeconomic factors.
4. To investigate whether the addition of psychosocial and socioeconomic factors to CVD risk **prediction** models can enhance their clinical performance.

1.6.2. Theoretical orientation and hypotheses formulation

After reading the literature and specifying my overall aims, I formed the following causal diagram of how socioeconomic/psychosocial factors could cause CVD, among those free from CVD at baseline (figure 14).

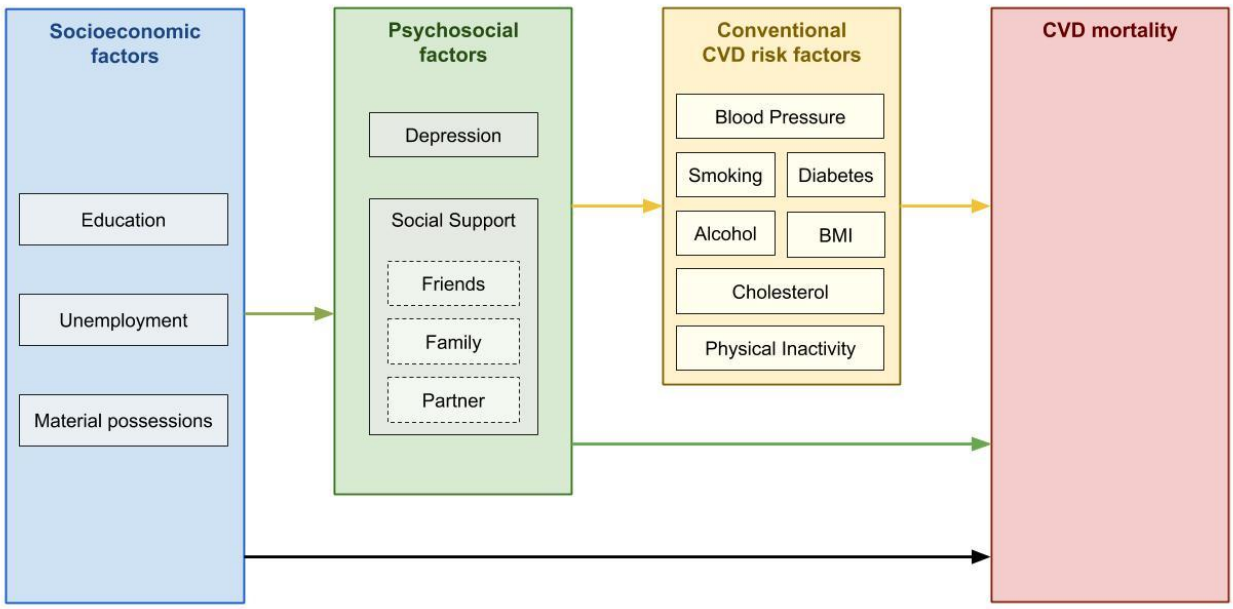


Figure 14. Theoretical direction of causal effects from three groups of risk factors, in influencing mortality from cardiovascular disease (CVD).

1.6.3. Specific objectives

1) **Causality using Mendelian randomization**
   a. To investigate the association between genetic predisposition towards education and subsequent risk of CHD, using two-sample Mendelian randomization (MR). This will inform the question of whether the association between education and CHD is causal or not.

   b. To further investigate the robustness and plausibility of my findings, additional analyses will be performed as:

      i. To evaluate whether the two-sample MR estimate obtained under 1a) could have been confounded by pleiotropy, using MR-Egger, Median-MR, Modal-MR, reverse-MR, as well as by triangulation of published results from twin studies and natural policy experiments.

      ii. To triangulate the results from 1a-b) with estimates of the observational association between education and:
         1. CHD prevalence, by using public data from NHANES.
         2. CHD incidence, by transforming published estimates from MORGAM.
         3. CHD incidence, by using unpublished data from HAPIEE.

      iii. To investigate the association between genetic predisposition towards education and subsequent changes to conventional CVD risk factors, using two-sample Mendelian randomization (MR).

2) Mediation

a. To investigate to what extent the observational association between socioeconomic factors and CVD might be mediated by conventional CVD risk factors and/or psychosocial factors.

b. To investigate to what extent the observational association between psychosocial factors and CVD might be mediated by conventional CVD risk factors and/or other psychosocial factors.

3) International differences

a. To investigate whether the inclusion of psychosocial and socioeconomic risk factors attenuates the additional cardiovascular hazard seen among participants from Russia, that is not seen among participants from Poland or the Czech Republic.

4) Prediction

a. To construct two new models that predict the incidence of CVD among healthy participants:
    i. A recalibrated SCORE: To recalibrate the existing SCORE model, using contemporary data from Eastern Europe.
    ii. An augmented SCORE: To extend the New SCORE model, by the addition of four psychosocial/socioeconomic predictors (education, marital status, unemployment, depression) and three behavioural factors (BMI, physical inactivity, antihypertensive use). In evaluating performance (internally), to specifically attend to reclassification around clinical intervention thresholds.

b. To externally validate both new models, using data from the Estonian Biobank.

2. Methods

2.1. Mendelian randomization

A range of study designs exist that offer additional insights into causal inference, beyond that of conventional epidemiological observational studies. These include natural policy experiments (and interrupted-time series designs) where exposure is suddenly modified at the population level. Regression discontinuity designs exploit the instances where social systems force a dichotomous threshold on a naturally continuous trait, such as forcing children with a biological age range of 12-months into annual educational cohorts. Studies of monozygotic twins can clarify whether observational associations might be attributable to unmeasured confounding from genetically inherited material, or parental influences shared by both twins. As an extension of this, siblings share only 50% of each other's genetic material but such studies can be many times larger, thereby increasing net statistical power. Finally, instrumental variables induce changes to exposure, but cannot have independent effects on the outcome. The randomization and allocation of treatment which occurs during a randomized controlled trial can be thought of as an instrument and is often seen as the gold standard. Examples of other instruments include random allocation into military service (as was done in the USA during the war against Vietnam), as well as the genetic instruments used in Mendelian randomization studies.

2.1.1. Background to Mendelian randomization

2.1.1.1. Principles of Mendelian randomization

Mendelian randomisation analysis uses genetic variants associated with a risk factor (e.g. education), to make causal inferences about how environmental changes to the same risk factor would alter the risk of disease (e.g. CHD).(224) By comparing the risk of disease across participants who have been grouped by their genotype, this enables the causal effect of a risk factor to be approximated with substantially less bias than that in a traditional observational analysis. Genetic markers of a risk factor are largely independent of confounders that may otherwise cause bias, as genetic variants are randomly allocated before birth.(225) This, as well as the non-modifiable nature of genetic variants, provides an analogy to trials, where exposure is allocated by random and is non-modifiable by subsequent disease.(225)

Finally, observational epidemiological analyses may be biased in cases where some participants show systematic reporting bias (when reporting either the exposure or outcome), which is different in magnitude when compared to the response bias shown by other participants. In Mendelian randomization analyses, both exposure and outcomes are proxied using genetic markers, which themselves are derived from GWAS analyses that best fit the average of the population sampled. The genetic markers themselves are objective, and do not suffer from the same problem of differential response bias. This is another example of where Mendelian randomization analyses may produce estimates that are closer to the true cause an effect.

2.1.1.2. Two-sample Mendelian randomization

Up until relatively recently, Mendelian randomisation (MR) analyses have been conducted on single datasets where data on genotype, risk factor and outcome were measured for all participants (known as “one-sample Mendelian randomisation”). However, performing advanced analyses on pleiotropy requires larger sample sizes in order to maintain statistical

power. This would require data pooling across dozens of studies, which is administratively difficult to organize. As an alternative, summary-level data from large genome-wide associations study (GWAS) consortia have become increasingly available in the public domain. It is possible to use such data to conduct Mendelian randomisation analyses, whereby gene-exposure measures are taken from one GWAS and gene-outcome measures are taken from another GWAS (altogether known as “two-sample Mendelian randomisation”).(226)

2.1.1.3. Sensitivity analyses for pleiotropy

Further methodological developments, including Mendelian randomisation-Egger (MR-Egger), weighted median MR, and mode-based methods, can all be employed as sensitivity analyses to additionally investigate any pleiotropic effects of the genetic variants (i.e., when genetic variants for education exert their influence on heart disease through an “off-target” pathway that bypasses the education phenotype (See Figure 15 for details).(226-228) The Mendelian randomisation method has successfully been applied to a range of biological and behavioural exposures.(229, 230) I am aware of just two studies that have applied it to investigate a socioeconomic exposure: a polygenic score for education has previously been associated with the development of myopia and dementia.(136, 137) However, these studies did not investigate the possibility of unbalanced horizontal pleiotropy.

Complex exposures, especially those instrumented by multiple SNPs of unclear biological role, may result in unbalanced horizontal pleiotropy. If this possibility is not examined, then such an omission might be considered to be the single most important weakness of such an MR study. Accordingly, I wanted to examine the potential for pleiotropy as rigorously as possible.

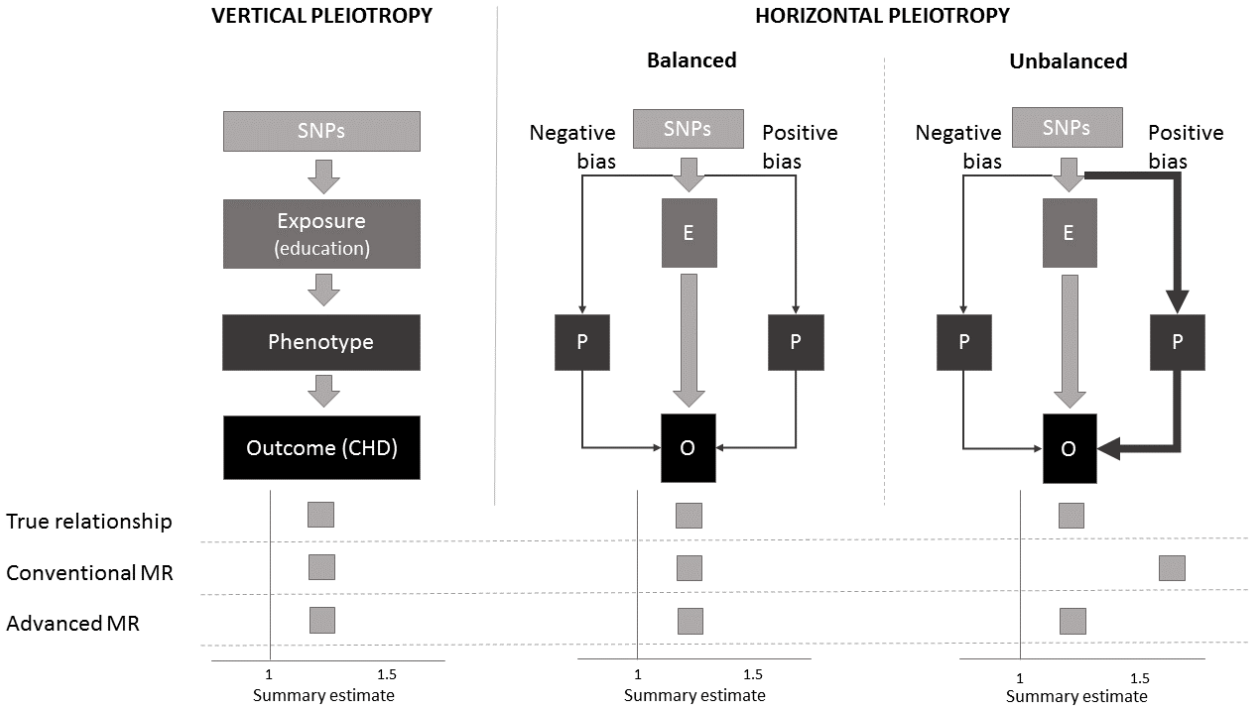


Figure 15. Theoretical illustration of pleiotropic phenomena on estimates derived from Mendelian randomization (MR) analyses.
In the case of vertical pleiotropy, the conventional MR assumptions are satisfied as the intermediating phenotype lies on a single causal pathway. In the case of horizontal pleiotropy, one or more phenotypes lie on a different parallel causal pathway. When the effects of the SNPs on the outcome through various parallel intermediating phenotypes are balanced, then estimates derived with conventional MR estimates should be valid. However, when the effects of SNPs on the outcome are systematically distorted towards one of the parallel alternative pathways (unbalanced horizontal pleiotropy), conventional MR estimates are invalid and biased due to confounding from the parallel pathway. Advanced MR techniques (representing techniques such as MR-Egger and weighted median MR), accounting for presence of unbalanced pleiotropy of the genetic instrument, should nonetheless produce an estimate that is closer to the true underlying association between the risk factor and outcome.
Figure adapted from White J. et al. JAMA Cardiol. 2016.(231)

2.1.2. Background to GWAS

In order to conduct a valid Mendelian randomization study, researchers need to identify a valid genetic instrument which induces changes to the exposure of interest (e.g. education). This is typically done by searching the literature of published genetic association studies.

Genome-wide association studies (GWAS) were originally often case-control studies in their nature, where a group of patients with the disease were compared against those without disease. Logistic regression (adjusted for age and sex) was used to test approximately one million genetic exposure (typically single-nucleotide polymorphisms [SNPs]), and identify those SNPs surpassing a standard Bonferroni adjustment (5×10^{-8}, also known as *genome-wide significant*).

In case that genome-wide significant associations are found, then some of these significant SNP may be located physically close to one another on the same genomic region of a chromosome in a process known as *genetic linkage*. Genetic linkage is one the reasons behind *linkage disequilibrium*, which denotes how two SNPs may not be inherited in a completely random fashion. In case that a GWAS identifies a cluster of correlated and physically co-located SNPs, such studies typically present just one 'lead SNP' in their final results. A process of *pruning* is performed to identify such lead SNPs, which ultimately selects the SNP with the smallest P-value as the lead SNP.

The central dogma of molecular biology states that alterations to DNA cause alterations to RNA which causes alterations to proteins (and subsequently, function and phenotype). Notably, alterations to the lead SNP may or may not cause downstream phenotypic change in the trait of interest. For example, it may be that the SNP with the largest P-value, identified as a 'lead SNP' is not causal for the trait of interest but is merely co-inherited with a causal SNP. Despite this nuance, people that carry non-causal lead SNPs are, on average, likely to also carry the unknown causal SNPs linked to the lead SNP, and are thereby likely to be randomly allocated to the trait of interest.

The case-control approach to a GWAS can also be modified to investigate traits which are distributed normally (such as BMI or education). In this case, linear regressions are used to predict the trait (e.g. education) as the outcome with age, sex, and each SNP being entered, one-by-one, in millions of univariate models. Some analyses additionally control for population stratification and additional cohort/period dummy variables.

During the start of my PhD (in May 2015), I performed a literature search for GWAS findings for any of the commonly measured socioeconomic and psychosocial traits. I was unable to find any study which had identified more than 5 SNPs which were subsequently replicated in an external cohort. Using less than 5 SNPs for an MR study is prone to weak-instrument bias, as well as eliminating the possibility of performing advanced tests for pleiotropy.

2.1.2.1. Details of GWAS on education

In May 2016, the Social Science Genetic Associations Consortium (SSGAC) published a GWAS of education that identified 74 independent lead SNPs, replicated in an external cohort. This was an important and unexpected finding for me, as this was the first time that so many independent SNPs had been identified for a socioeconomic/psychosocial trait. This created the potential to investigate in more detail whether education causes a common health outcome, such as coronary heart disease.

The GWAS of education was conducted by meta-analysing summary-level data from 64 cohorts (each with sample sizes ranging from n=318 to n=76,155), altogether on 293,723 participants. Participants were men and women of a wide age range, of European ancestry. All cohorts were recruited in either European countries, USA and Australia.

Missing data in SNP-education associations

Most common genotyping chips measure between 0.2 to 2 million SNPs. Okbay et al wanted to evaluate associations across almost the entire genome, i.e. 9,256,490 SNPs. While it may be unreasonable to impute 90% for non-genetic studies, this is less problematic in genetic studies, as the degree of correlation between individual SNPs (i.e. linkage disequilibrium) is so high.

Each of the 64 cohorts imputed their missing genotype data with reference to some version of the 1000 Genomes project. Association analyses were performed within each cohort on all genotyped and imputed SNPs. and summary statistics were meta-analysed by Aysu Okbay. Prior to meta-analysis, a set of quality control (QC) filters were applied to cohort-level association results to make sure they were statistically reliable. SNPs with low imputation accuracy were filtered out as part of QC. The actual imputation accuracy thresholds used to filter out badly imputed SNPs differed by the type of imputation software used by the underlying cohort: 0.6 for MaCH, 0.7 for IMPUTE2, and 0.8 for PLINK.

As a result of differences in genotyping chips and imputation reference panels (different versions of the 1000 Genomes reference panel, or in a handful of cohorts, other sequence data comprising a similar set of SNPs) used by cohorts, and the quality control filters applied to cohort-level results, a slightly different set of SNPs from each cohort entered the meta-analysis. Hence, some SNP-exposure estimates were derived from larger samples while other SNP-exposure estimates were derived from smaller samples (max 349,306 participants, in the sample which did not overlap with my outcome dataset. Median SNP=347,477 participants. Least available SNP= 220,565 participants). In case that GWAS results were available for the full sample (349,306 participants), I report this below as “100% available / complete”. For the vast majority of SNPs (141 out of 162 SNPs, i.e. 87% of SNPs), data were non-missing for ≥90% of participants.

For an average SNP, GWAS results were missing for just 3.5% of participants. The distribution of missing data across SNPs ranked by their degree of missingness is shown in figure 16.

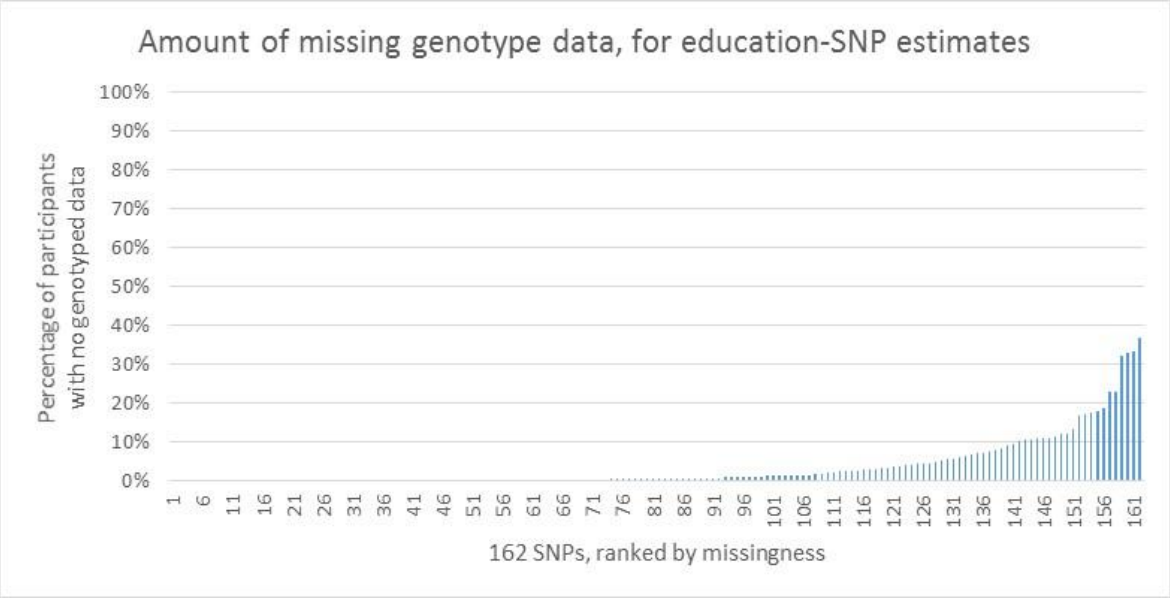


Figure 16. Amount of missing data for each of the 162 instrumented educational SNPs.

Overall, data completeness in SNP-exposure associations appears to be very good. Additional details about these SNPs, including details on missing data, are provided in “Data supplements.xls”, on the website of the journal where this research was published: www.bmj.com/content/358/bmj.j3542

Biological function

At current, it is unknown how these 74 SNPs (if causal) might exert their influence on education. Most of these loci are not known genes, making it possible that they may regulate the function of other unknown genes. Still, Okbay et al. went into considerable detail to investigate biology using other approaches, some of which is relevant to the question of whether these 74 SNPs may have pleiotropic effects on cardiac function in a pathway that bypasses education. To paraphrase their key findings [my additions in square brackets]:

"SNPs in regions that are DNase I hypersensitive in the foetal brain are more likely to be associated with education by a factor of ~5. Moreover, the 15% of SNPs residing in regions associated with histones marked in the central nervous system explain 44% of the heritable variation. This enrichment factor of ~3 for CNS is greater than that of any of the other nine [non-neurological] tissue categories in this analysis [such as cardiac tissue]. Given that our findings disproportionately implicate SNPs in regions regulating brain-specific gene expression, we examined whether genes located near education associated SNPs show elevated expression in neural tissue. Remarkably, the 13 GTEx tissues that are components of the CNS—and only those 13 tissues—show significantly elevated expression levels of genes. [i.e. no expression in non-Central Nervous System tissues].

The resulting 34 [gene] clusters paint a coherent picture, with many corresponding to stages of neural development: the proliferation of neural progenitor cells and their specialization, the migration of new neurons to the different layers of the cortex (forebrain development, abnormal cerebral cortex morphology), the projection of axons from neurons to their signalling targets (axonogenesis), the sprouting of dendrites, and neuronal signalling and synaptic plasticity throughout the lifespan (voltage-gated calcium channel complex)."

Altogether, this suggests that these 74 SNPs are primarily involved in neuronal development and are unlikely to directly (by alternative parallel pathways) cause changes to cardiac or vascular tissues.

2.1.2.2. Details of GWAS on CHD

To conduct two-sample MR, I needed to access summary-level data on the handful of top SNPs associated with education, and see if they are also associated with heart disease. For this, I primarily used data from the 2013 release of the Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics) consortium. They meta-analysed data from 39 studies, to look at genetic differences between 63,746 CHD cases against 130,681 controls. Most of these cases were relatively early onset (the typical age of presentation was between 50 to 60), with predominantly non-fatal as opposed to fatal cases.

2.1.2.3. Details of GWAS on 10 CVD risk factors

To look at causal effects from education where the outcome is not disease, but a CVD risk factor, I used data on 10 risk factors from 6 GWAS studies, described in table 3.

Consor-tium	Outcome	Parti-cipants	Website for data download
TAGC	Smoking	74,053	www.med.unc.edu/pgc/results-and-downloads
ICBP	Blood pressure	74,064	www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000585.v1.p1 \
GLGC	LDL-, HDL-cholesterol, and triglycerides	188,577	csg.sph.umich.edu/abecasis/public/lipids2013/
DIAG-RAM	Type 2 diabetes	149,821	diagram-consortium.org
MAGIC	Glucose	133,010	www.magicinvestigators.org
GIANT	BMI, height	339,224	portals.broadinstitute.org/collaboration/giant/

Table 3 Identification of potential Mediators (education to conventional CVD risk factors)

TAGC, Tobacco and Genetics Consortium. ICBP, International Consortium for Blood Pressure. GLGC, Global Lipids Genetic Consortium. DIAGRAM, DIAbetes Genetics Replication And Metaanalysis. MAGIC, Meta-Analyses of Glucose and Insulin-related traits Consortium. GIANT, Genetic Investigation of Anthropometric Traits.

2.1.3. Overview of the Mendelian randomization analyses performed

Prior to conducting an MR analysis, it is sometimes useful to check whether the exposure and outcome traits of interest share some common underlying genetic architecture. To investigate the genetic correlation between education and CHD, I used the "Lookup Center" function of the LD Hub platform (http://ldsc.broadinstitute.org).(232) On 28th October 2016, I downloaded an XLS file of genetic correlations, based on latest data that has been previously uploaded onto the LD Hub platform. Analyses were done by Linkage Disequilibrium Score Regression. After integrating two GWAS datasets and examining millions of SNPs across the entire genome, there was strong evidence for a negative genetic correlation between education and CHD (r_g=-0.324; r_g^2=0.104; P-value=2.1 x 10^{-12}).(233) To interpret this, educational outcomes can vary due to genetic and non-genetic variance. Within the domain of genetic variance, approximately 10% of it appears to be shared with the genetic variance of CHD (and vice-versa), whereby this correlation is negative. This correlation can arise for various reasons, so I next conducted Mendelian randomisation analyses to investigate the presence and direction of any causal effects.

My study of causality was composed of four research questions, indicated in blue in figure 17.

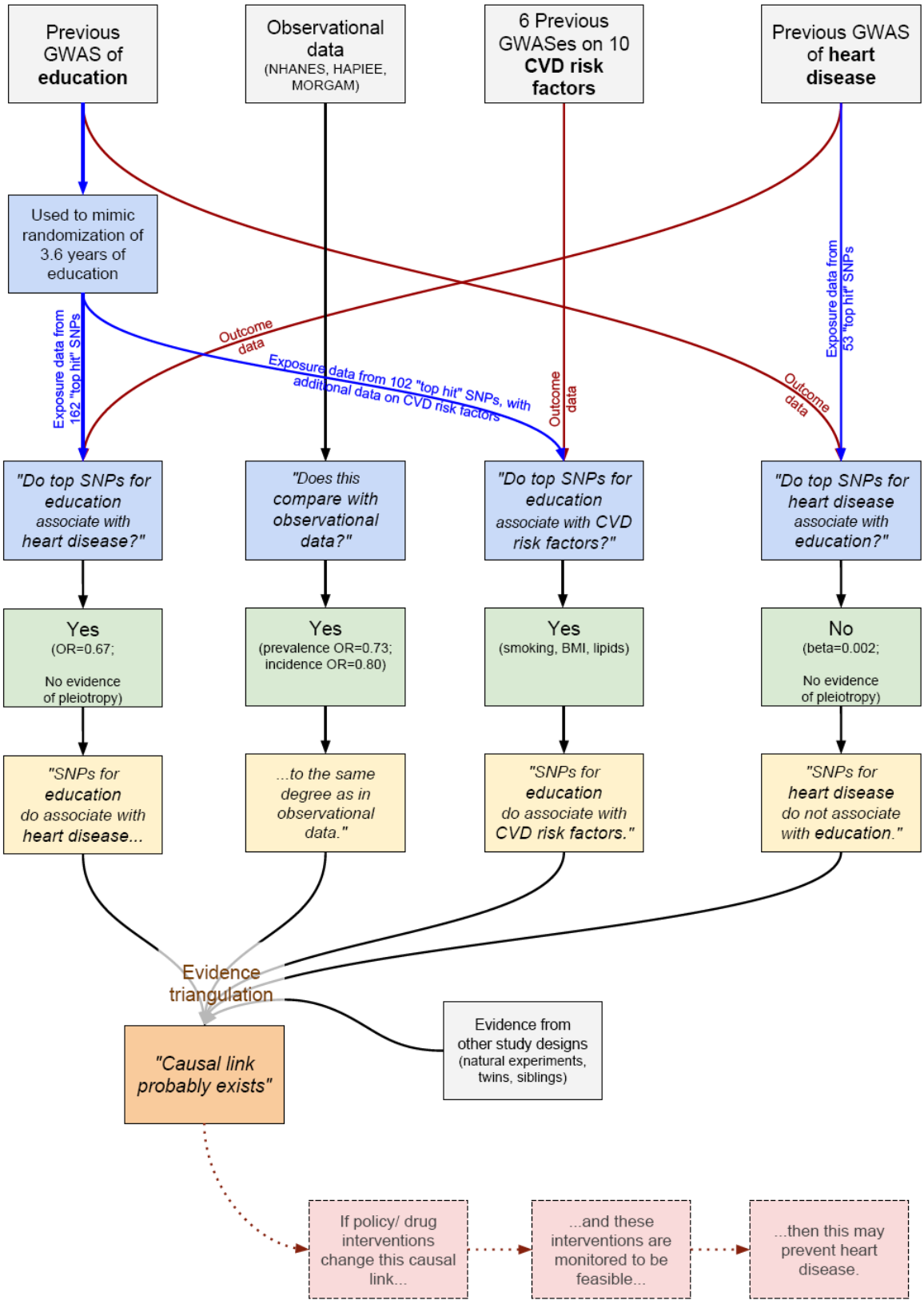


Figure 17. Overview of the main steps in this study.
Existing datasets (grey), hypothesis formulation (blue), key findings (green), their interpretation (yellow), conclusion (orange) and final suggestions for discussion (red).

Throughout all analyses, education was initially coded as “highest educational qualification”. Using the International Standard Classification of Education 1997 classification system (see supplementary table 1.3 of the original GWAS study(234), this was transformed into the number of typical years spent in education, to obtain that qualification (using country-specific conversion guidelines). Following harmonization, self-reported educational attainment was modelled linearly, expressed as one standard deviation (i.e. 3.6 years) of additional schooling. In this form, one year of vocational education was equivalent to one year of academic education, and I did not assume any qualitative differences in the *type* of education. CHD was defined as a composite of myocardial infarction, acute coronary syndrome, chronic stable angina or coronary stenosis of >50%, or coronary death.

2.1.4. Mendelian randomization data sources (exposure to education)

I will now describe in more detail the steps taken for question 1 (which additionally informs the steps taken for question 3, as illustrated in Figure 18.

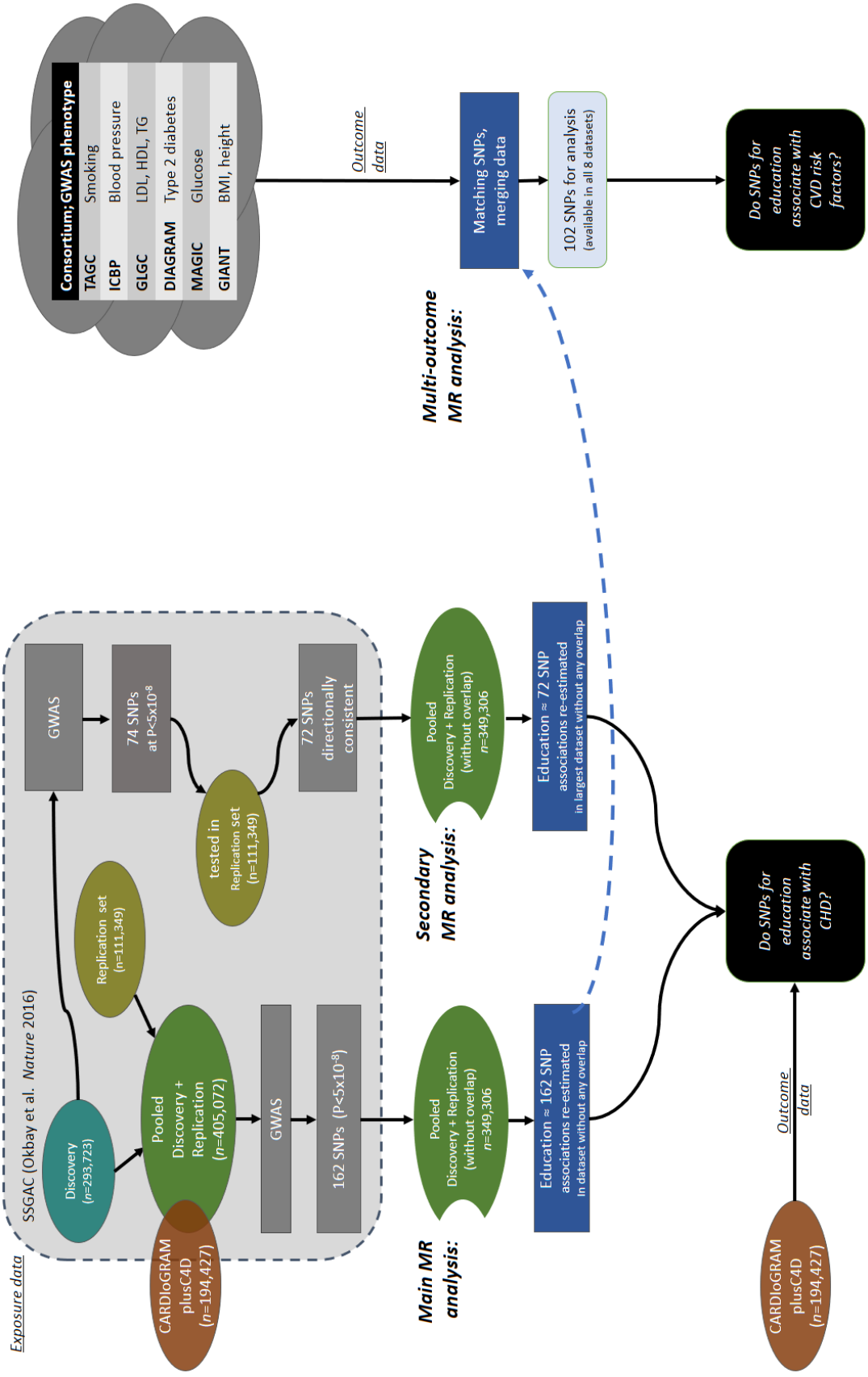


Figure 18. Flowchart of the derivation of two sets of education SNPs (162 and 72 SNPs), used in Mendelian randomization estimates from education.

The Main MR analysis was performed on a slightly larger set of SNPs than what was discovered and replicated in the original GWAS of education. Namely, I used 162 independent SNPs associated ($P<5.10^{-8}$; linkage disequilibrium $r^2<0.1$) with education in a meta-analysis of the discovery (Social Science Genetic Association Consortium) and replication (UK Biobank) datasets combined. This has the disadvantage that these 162 have not been validated in an external dataset, making it possible that they have more type1 errors and are hence not sufficiently predictive of education in external datasets (such as the outcome datasets); and they may be more pleiotropic for other traits. However, if there are few type-1 errors, then they can form a more powerful genetic instrument, which is particularly useful for sensitivity analyses of pleiotropy. Altogether, these 162 SNPs explained 1.8% of the variance in education in this discovery stage. This is sufficient to generate a strong genetic instrument with which to derive precise causal estimates (see Table 4 for power calculations).

As a sensitivity analysis, I repeated all analyses in a second set of MR analyses, where I used another a smaller set of 72 independent SNPs (at $r^2<0.1$) that were associated with education in the smaller discovery dataset (SSGAC) alone (293,723 participants; $P<5.10^{-8}$) and which subsequently tested to be directionally consistent in an independent replication dataset (UK Biobank). Ultimately, I decided to use the larger set of instruments (with 162 SNPs) in the main analysis instead of the smaller set of instruments (with 72 SNPs) in order to maintain sufficient statistical power for my sensitivity analyses.

To avoid biases due to overlapping datasets (i.e. where data from the gene-education association and gene-CHD association are derived from the same samples, and which can lead to bias in the direction of the observational association in the presence of a weak instrument),(235) excluded data from the following studies from the SNP-education data source: deCODE, WTCCC, KORA, THISEAS, and 23andMe, by contacting the authors of the original GWAS study.(234) The SNP-education estimates from this restricted dataset (n= 349,306) were highly correlated with the SNP-education estimates from the complete SSGAC dataset (Pearson's r for 162 SNPs=0.96 [p-value<0.001] and Pearson's r for 72 SNPs=0.92 [p-value<0.001]) (Figures 19 & 20 on subsequent pages), and were hence used in subsequent MR analyses.

	Genetic instrument of education	
	1st set of SNPs (162 SNPs)	**2nd set of SNPs (72 SNPs)**
R-squared (of variance in educational phenotype)	0.018	0.008
Actual N (CARDIoGRAM plusC4D)	194,427	194,427
Proportion of cases (CARDIoGRAM plusC4D)	0.327	0.327
Observational OR	0.8	0.8
N required for 80% power	42,832	96,372
Power at actual N	>0.99	0.98

Table 4. Power for conventional Mendelian randomization analysis (two-sided α=0·05). *Based on the method developed by Brion et al.(236).*

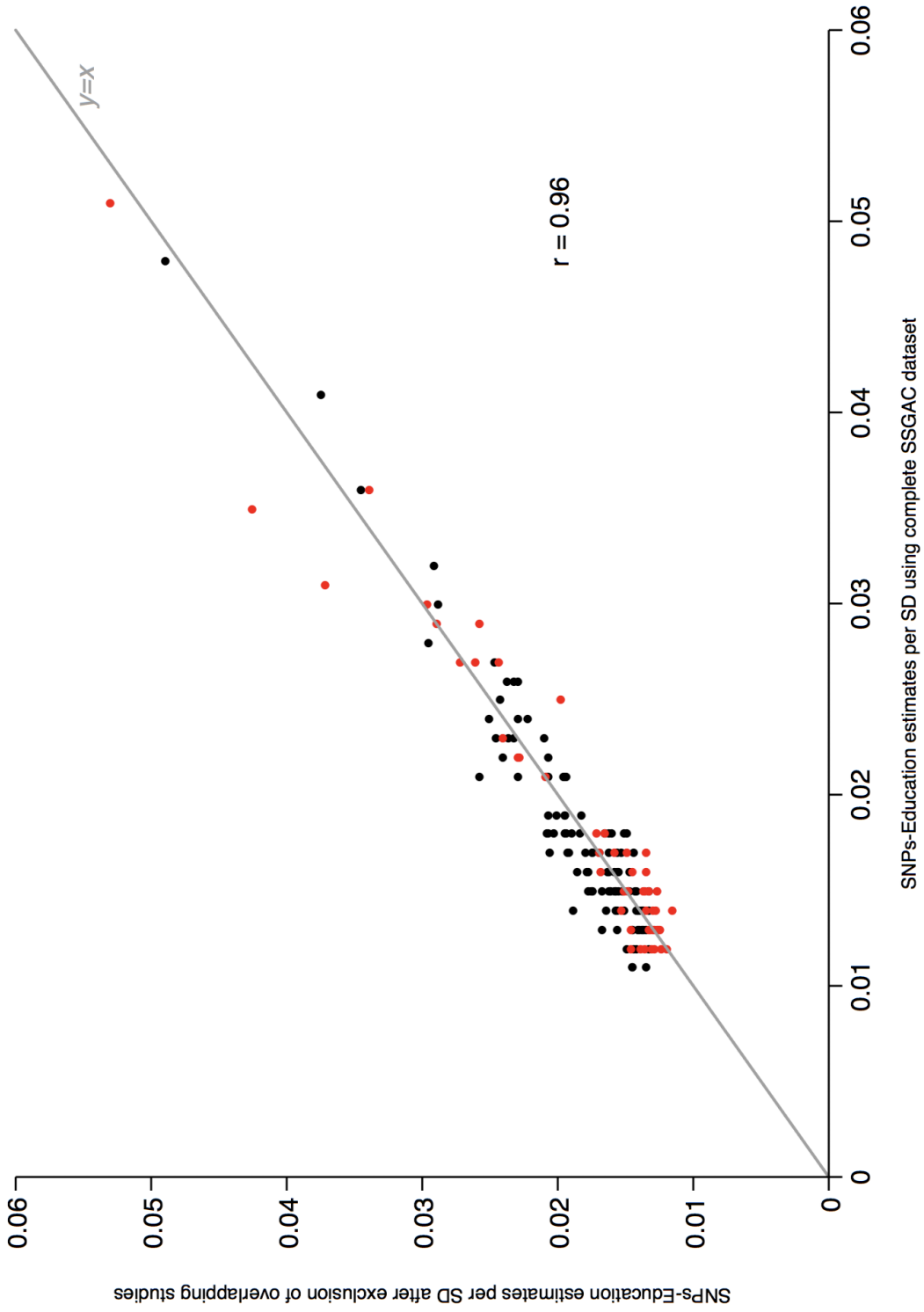


Figure 19. Scatter plot of SNP-education association estimates, comparing complete dataset vs. non-overlapping dataset (n=162 SNPs). *The complete SSGAC dataset (x-axis) was based on 405'072 participants. The restricted dataset (y-axis) was based on 349,306 participants without sample overlap. Median standard errors (SE) in the complete SSGAC and restricted dataset were both of 0.003. Red points indicate 51 SNPs that remained associated with the education phenotype above the GWA threshold ($P>5 \times 10^{-8}$) in the smaller dataset without sample overlap.*

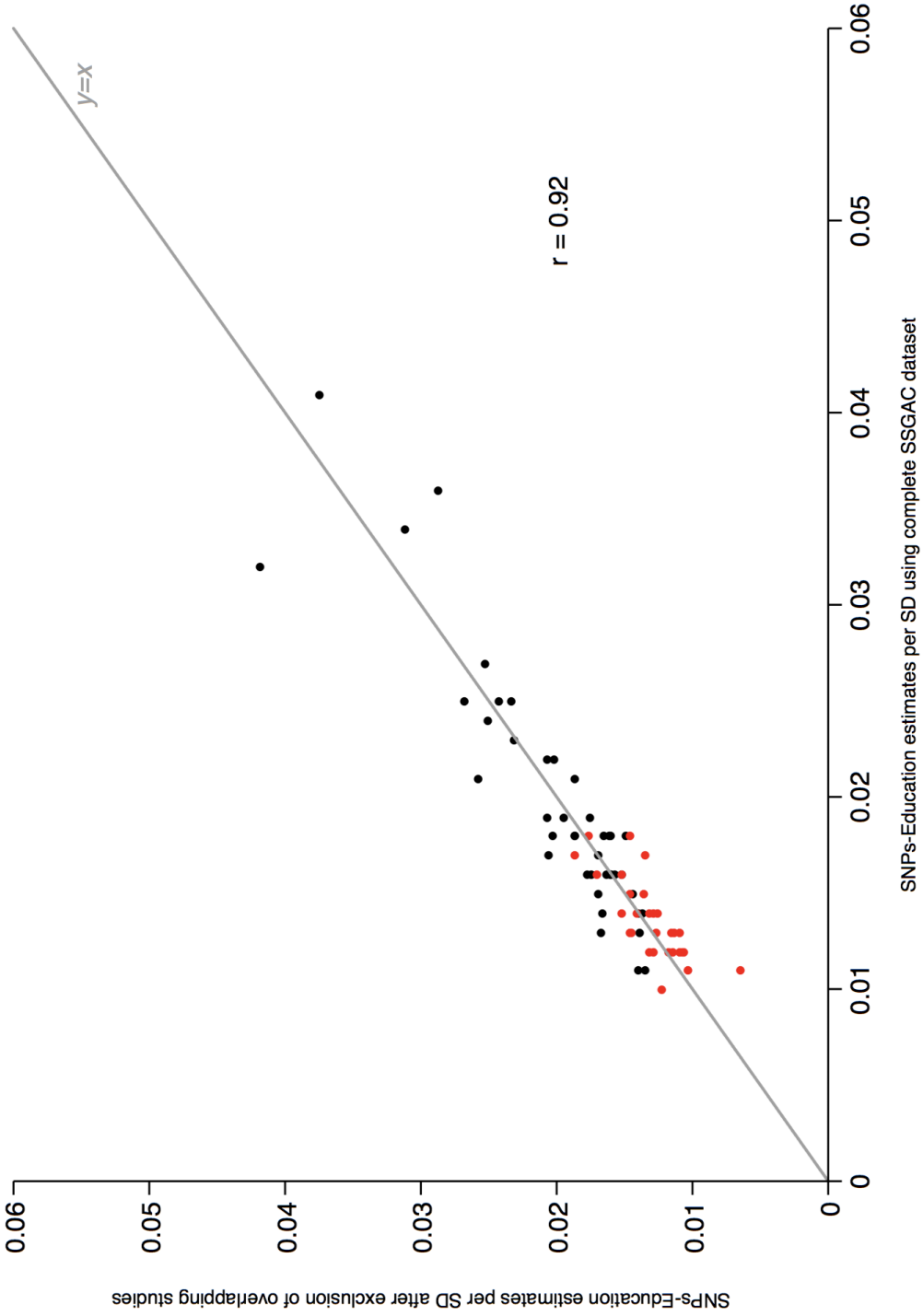


Figure 20. Scatter plot of SNP-education association estimates, comparing complete dataset vs. non-overlapping dataset (n=72 SNPs). *The complete SSGAC dataset (x-axis) was based on 293'723 participants. The restricted dataset (y-axis) was based on 349,306 participants without sample overlap (and where the replication data was additionally used, to allow for most precise estimates while concurrently minimizing bias). Median SE in the complete SSGAC and restricted data dataset were both of 0.003. Red points indicate 29 SNPs that remained associated with the education phenotype above the GWA threshold (P>5 x 10^{-8}) in the dataset without sample overlap.*

2.1.5. Mendelian randomization data sources (CHD outcome)

Once the instruments had been selected and the strength of association between SNP-exposure (i.e. beta coefficients) had been determined, I looked to see if the same SNPs are associated with the outcome of interest, namely CHD. The CARDIoGRAMplusC4D dataset I used (237) is actually a collection of 4 datasets (genotyped CARDIoGRAM, genotyped C4D, genotyped Metabochip, Imputed data). I retrieved summary-level data for either the same SNP (115 out of 162 SNPs) or for a proxy SNP in high linkage disequilibrium (47 out of 162 SNPs at r2>0.8) following a flowchart of preferential data quality as detailed in Figure 21. Along similar lines, outcome data for the secondary set of 72 instruments is presented in figure 22.

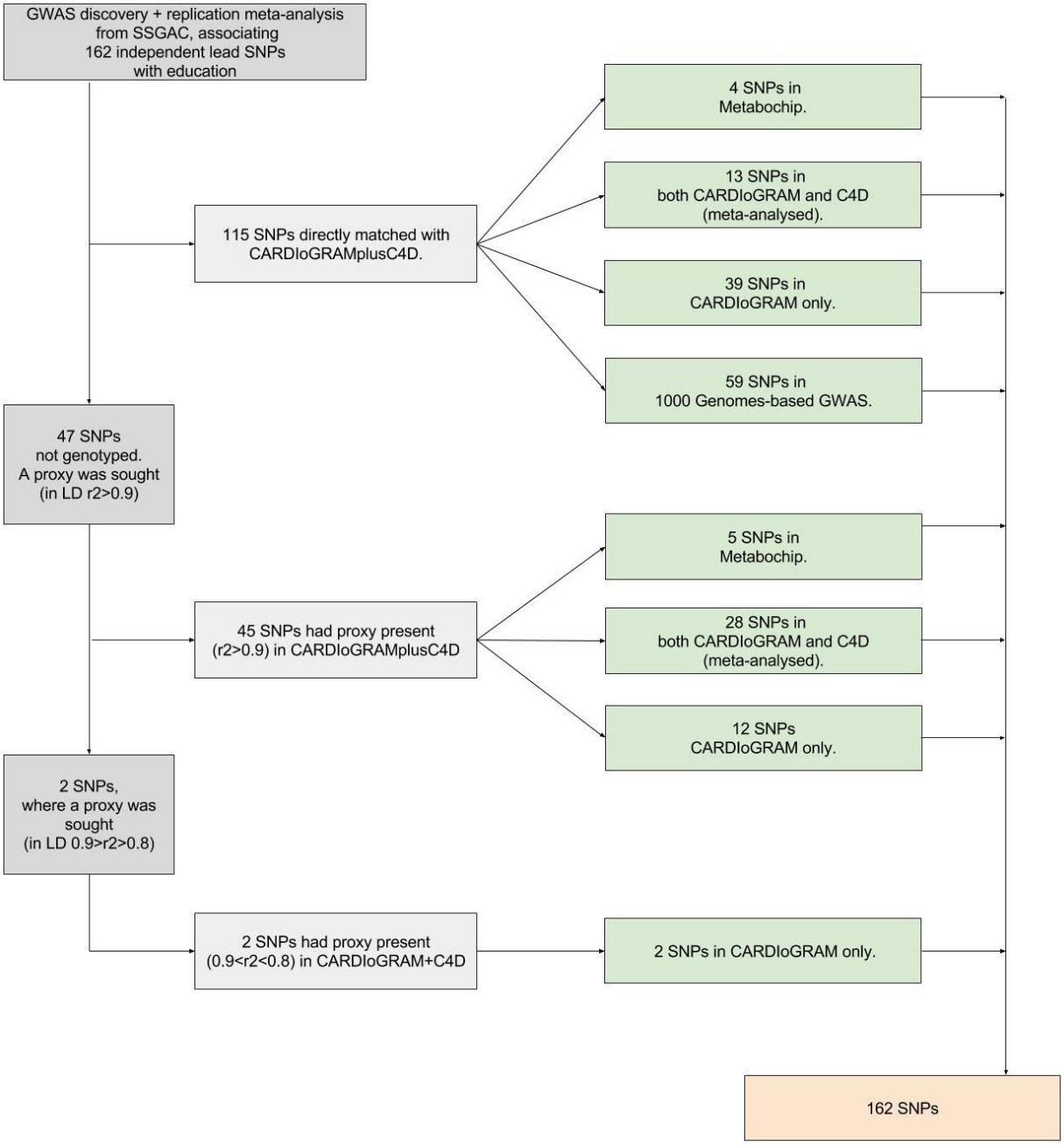


Figure 21. Flowchart illustrating how SNPs identified in the SSGAC education GWAS were mapped against SNPs reported in CARDIoGRAMplusC4D (n=162 SNPs). *Where necessary, proxies were retrieved using the SNP Annotation and Proxy Search online tool (SNAP, http://archive.broadinstitute.org/mpg/snap/ldsearch.php ; reference panel = 1000 Genomes; LD threshold r2>0.80).*

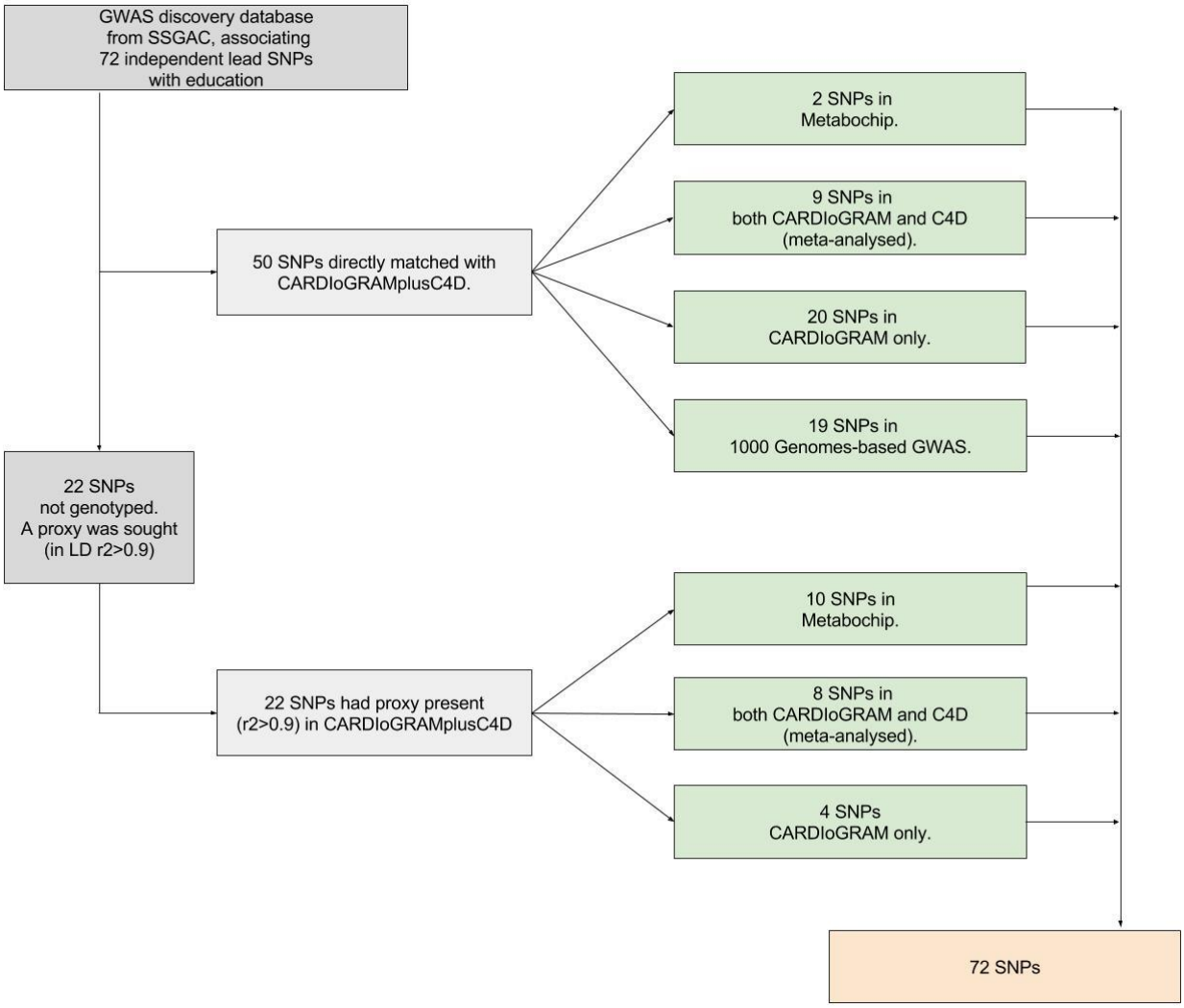


Figure 22 Flowchart illustrating how SNPs identified by SSGAC education GWAS were mapped against SNPs listed in CARDIoGRAMplusC4D (n=72 SNPs). *Where necessary, proxies were retrieved using the SNP Annotation and Proxy Search online tool (SNAP, http://archive.broadinstitute.org/mpg/snap/ldsearch.php ; reference panel = 1000 Genomes; LD threshold r2>0.80).*

2.1.6. Statistical analysis (main Mendelian randomization estimate)

For all MR analyses, alleles from SSGAC and CARDIoGRAMplusC4D datasets were aligned to correspond to an increase in educational attainment. This step, if not performed (as is done in some of published papers) invalidates some advanced analyses such as MR-Egger, discussed later.

Conventional Mendelian randomization

I used conventional (also termed 'inverse variance weighted') Mendelian randomisation analyses, by regressing the SNP-education associations against the SNP-CHD associations (outcome), where each SNP was one data point. This regression was weighted by the standard error of each SNP (whereby standard errors are larger for rarer genetic variants, and also larger if measured on a smaller subset of participants). Contrary to MR-Egger, the conventional MR regression line was forced to pass through the origin.(227)

2.1.7. Statistical analysis (sensitivity Mendelian randomization)

A range of sensitivity analyses were used to investigate to what degree the presence of unbalanced horizontal pleiotropy might bias the result of conventional Mendelian randomisation. These methods allow some of the Mendelian randomisation assumptions to be relaxed, and replaced with different and somewhat orthogonal assumptions. Consistency of results across a range of methods that make different assumptions regarding pleiotropy strengthens causal inference, while divergent results may indicate that genetic pleiotropy is biasing some of these results.

2.1.7.1. Penalized weighted median Mendelian randomization

A penalized weighted median MR analysis was conducted (implemented in Stata using the *mrrobust* package; available at: https://github.com/remlapmot/mrrobust).(226) This gives more weight to genetic variants close to the median causal estimate. Weighted Median methods yield robust and precise results even when up to 50% of the weight in the analysis stems from invalid genetic variants.(226)

2.1.7.2. Standard MR-Egger regression

MR-Egger regression was applied as described by Bowden et al.(227) Based on the same principles as the Egger test (which assesses small study bias in meta-analysis) the method is similar to conventional MR analyses. However, the regression is not constrained to pass through the origin. A significant departure of the y-intercept from the origin gives evidence for the presence of unbalanced pleiotropy. If the level of pleiotropy is independent of the strength of the association between SNPs and education, the MR-Egger estimate thus represents the true causal effect, even if all the genetic variants present pleiotropic effects (as per the InSIDE rule).(227) The standard error (SE) of the causal estimate was corrected by dividing the reported SE of the estimate by the residual SE.

2.1.7.3. Adjusted MR-Egger (+SIMEX)

All MR approaches rely on the fact that *the SNP-exposure association is true (NO Measurement Error [NOME] assumption)*. Whenever the SNP-exposure association estimates are spurious (violation of the NOME assumption), weak instrument bias can distort the causal effect estimate (specifically diluting it towards the null value).(238) Using the I2 statistic, I quantified the expected dilution in the MR-Egger causal effect estimates due to the variance of the estimates of the SNP-education association: I2 was only moderate for the set of 162 SNPs (I2=66%; potential dilution of 44%), whereas I2 for the set of 72 SNPs indicated a reduced risk of bias (I2=93%; potential dilution of 7%). As described by Bowden et al, I applied simulation extrapolation (SIMEX; implemented in R using the *simex*

package) to adjust the MR-Egger causal estimates to account for violations to the NOME assumption, using 10,000 simulations (238). Simulation extrapolation analyses were conducted using R v3.3.1, with the assistance of Julien Vaucher and Jack Bowden.

2.1.7.4. Mode Based Methods (assuming Zero Modal Pleiotropy)

Recently developed methods relax the conventional MR assumptions, and instead form a less stringent assumption of *Zero Modal Pleiotropy.(239)* This postulates that pleiotropic SNPs are unlikely to converge on the same modal (most common) estimate due to pleiotropic effects not being identical. In contrast, valid SNPs are more likely to converge on the same, common modal estimate. I performed three analyses to exploit this assumption (with the assistance of Julien Vaucher, Fernando Pires Hartwig and Jack Bowden).

a) <u>Mode-Based Estimate</u>

   In my first analysis, I used the Mode-Based Estimate (MBE). With an infinite sample (i.e. no measurement error), the MBE would use the modal causal estimate (i.e. the most common instrumental variable estimate out of the 162 SNPs, where the instrumental variable estimate for each SNP is derived by dividing the SNP to CHD estimate by the SNP to education estimate). In finite samples, the MBE uses the mode of the smoothed empirical density function of causal estimates (where the instrumental variable estimate for each SNP is up weighted by its relative precision, in comparison to other SNPs). A tuning parameter φ regulates the bias-variance trade-off. I explored a range of these, following which φ=0.5 was chosen to best fit the data. The analysis makes the assumption that the most commonly observed causal effect estimate comes from valid genetic instruments, and it can provide a consistent causal effect estimate even if the majority of (non-modal) genetic instruments are invalid.

One advantage of the Mode-Based Estimate is that it is less influenced by outlying (and possibly pleiotropic) genetic instruments without formally removing them from the analyses, thus making full use of the data. However, the uncertainty around the point estimate can sometimes be prohibitively wide. For this reason, I exploited the *Zero Modal Pleiotropy* assumption using another strategy, which involves actually removing some genetic instruments from the analysis.

b) Largest Homogeneous Subset-MR

In my second modal analysis, SNPs were removed, one-by-one, until the final set of SNPs contained only sufficiently similar (according to some criteria) effect estimates. As such this final set of SNPs can be thought of as a relatively "homogeneous subset". The steps I took are:

1. Begin with the set of 162 SNPs. Evaluate the heterogeneity among each of the 162 causal estimates, using *Cochran's Q* statistic.(240)
2. Remove the SNP which contributes most to heterogeneity.
3. Repeat the proves, until a P-value threshold of heterogeneity is reached (e.g. P>0.05). Smaller P-values denote more heterogeneity (close to the original set of 162 SNPs) while larger P values include fewer SNPs and are hence more stringent.

c) Largest Homogeneous Subset-MR: removing *most causal* variants

My third analysis was similar to the second one described above. This time, instead of removing any SNP (either at the left- or right-hand tail of the causal effect distribution) I only removed those SNPs that provided the strongest causal estimates (on one side of the tail), until the heterogeneity P value was attained as above. This method is unlikely to provide a valid causal estimate, as its result will be biased towards the null. However, it is an extreme example of a very stringent sensitivity check which asks the question *"What if the most outlying SNPs, all of which produce the strongest causal estimates, were*

deemed as invalid (due to having suspected horizontal, unbalanced pleiotropic effects on CHD)?"

2.1.7.5. Reverse direction MR

To check for whether genetic risk for coronary events might be a causal factor for educational attainment (my fourth research question), I performed Mendelian randomisation in the opposite direction (so-called bi-directional Mendelian randomisation) using 53 SNPs associated with CHD. Under conditions of massive pleiotropy, genetic risk of coronary events might also predict educational outcomes.

I used data from the CARDIoGRAMplusC4D Consortium to extract SNP-CHD estimates for 53 independent SNPs (at r2<0.02) that were GWAS significant ($P<5x10^{-8}$).(237, 241-243). I directly matched these with the corresponding SNPs from the SSGAC GWAS, involving 328,917 individuals (**Supplementary Dataset 3**).(234) This analysis was performed using data where the underlying participants overlapped slightly between the SNP-exposure and SNP-outcome estimates. However, as such overlap biases results away from the null,(244) and since my finding was quite definitively null, I did not purse seeking non-overlapping data in this instance, for this post-hoc sensitivity analysis. The causal MR analyses were conducted similar to the analyses described above, including the main sensitivity analyses.

2.1.7.6. From education to CVD risk factors

For my third research question, I applied conventional Mendelian randomisation to investigate whether genetic predisposition towards longer education could lead to improvements in the established cardiovascular risk factors. In this analysis, I discarded 60 SNPs with missing data on one of the cardiovascular risk factors, and thus used a smaller set of 102 SNPs.

2.1.8. Comparison with observational data

Once I had assembled a range of causal estimates, it is useful to compare this against a similar estimate derived from conventional observational epidemiological methods (e.g. estimate of incidence from cohort studies, and/or cross-sectional prevalence from case-control studies). Since the MR estimates are likely to be less precise than observational estimates, such direct comparisons can assist in making inferences about whether the observational estimates are likely to be biased upwards, downwards or neither. Importantly, the exposure and outcome should be defined in similar ways in observational data, as I did in my causal analysis.

After reviewing the literature of observational epidemiological findings in this area, I concluded that the vast majority of studies used a substantially different definition of exposure. Virtually all the time, an ordinal categorical definition was used in observational studies, as opposed to the continuous trait used in my causal analysis. Moreover, the number of categories varies substantially from around two to five categories. One recent meta-analysis summarized this evidence by comparing estimates of “least educated” against “most educated categories”, which might artificially increase between-study heterogeneity simply from the range of categories applied in various studies.(8)

I decided to provide my own observational estimates, using a mixture of to provide estimates from a case-control design, plus two methods to provide estimates from a prospective cohort design.

2.1.8.1. Observational case-control data

For prevalent CHD cases in cross-sectional data, I analysed 43,611 participants (1,933 cases) from 8 baseline samples of the National Health and Nutrition Examination Surveys (NHANES), from an instantly downloadable website www.cdc.gov/nchs/nhanes/ .(160) Figure 23 illustrates the sample selection process.

Just as in the causal analysis, I transformed the self-reported highest qualification into number of years in education, using the ISCED 1997 system. The association between education and self-reported CHD prevalence was calculated using multivariable logistic regression adjusting for age and sex, and with their recommended weighted to account for non-random sampling, response bias and geographical clustering.(160)

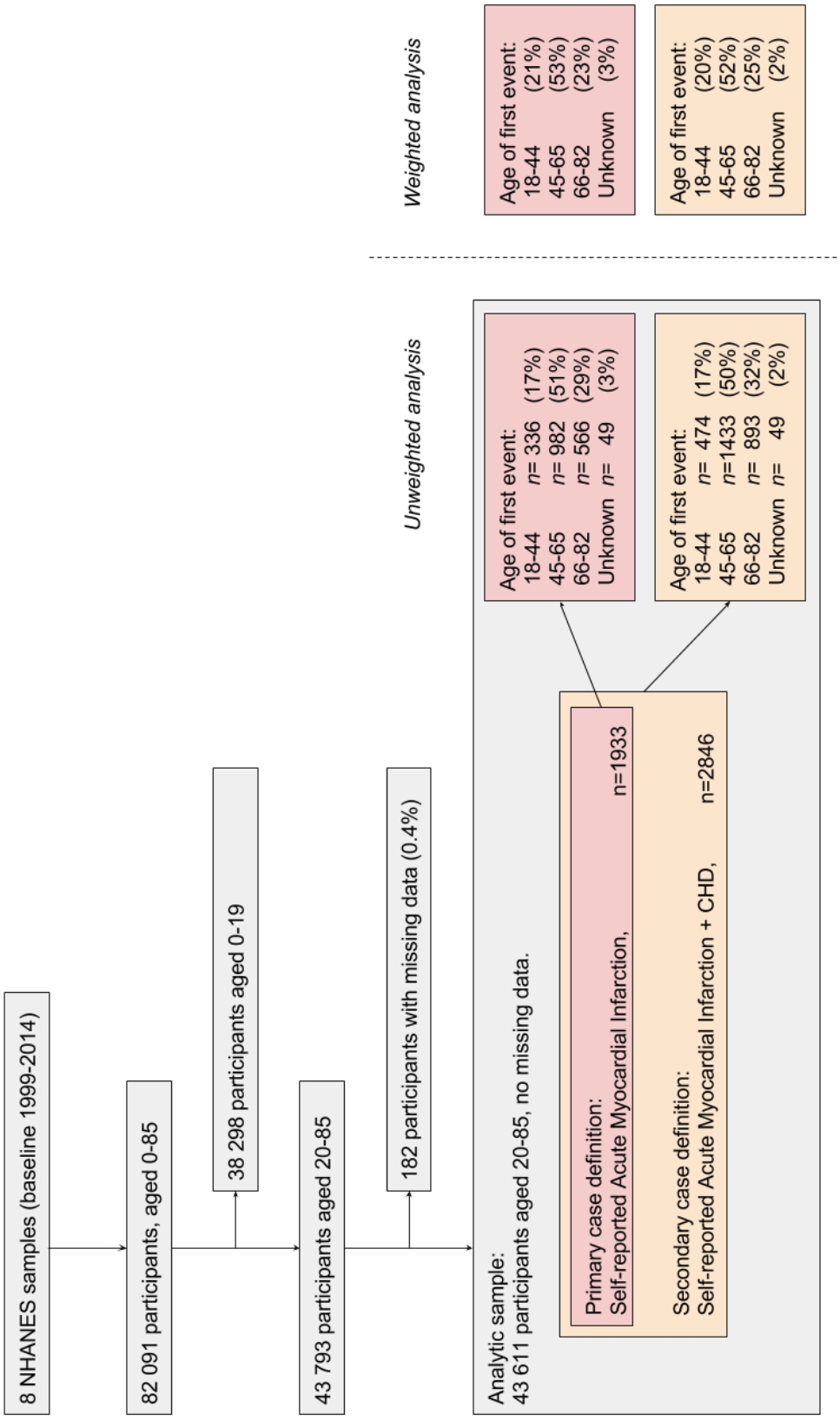


Figure 23. Flowchart of participants in observational NHANES analysis.

2.1.8.2. Observational prospective data

The association between education and the incidence of clinically-verified CHD was derived from two data sources. First, I first analysed 23,511 participants (632 cases) from the Health, Alcohol and Psychosocial factors In Eastern Europe (HAPIEE) study,(245) using a Cox proportional hazard regression adjusting for age, sex, and country.(245) Further details of the HAPIEE study are shortly provided in Chapter 2.2.

Second, I reanalysed published summary-level data from 97,048 participants (6,522 cases) of the MOnica Risk, Genetics, Archiving and Monograph (MORGAM) study in Europe. Here, the published hazard ratios (HR) were initially expressed in terms of country-specific top vs. bottom tertiles of education.(246) I took the log of the hazard ratios corresponding to the lowest and highest educational tertiles, and divided it by the number of years of education between the lowest and highest tertiles from another publication,(12) and inverted the sign (to express the estimate per 1-year of additional education). Results were multiplied by 1-SD (3.6 years of education), and log HRs were back transformed into HR of CHD per 1-SD increase in education. Country and sex-specific MORGAM estimates were then pooled using fixed-effects inverse-variance meta-analysis. HAPIEE and MORGAM estimates were meta-analysed similarly, to derive a summary estimate of incidence. These observational analyses are summarized in Table 5.

Since my MR analyses assumed a linear relationship between education and CHD, I additionally used the individual-level data from NHANES and HAPIEE to model the dose-response observational association between years of education (as an ordinal categorical variable) and risk of CHD.

Study	NHANES (1)	HAPIEE (2)	MORGAM (3)
Access policy	Public domain, with all required data instantly downloadable	Access upon application to principal investigators.	Access upon application to principal investigators.
		My analysis has not been previously published.	My analysis is based on previously published work.
Design	Cross-sectional	Longitudinal	Longitudinal
Country	United States of America	Russia, Czech Republic, Poland, Lithuania	9 European countries.
Baseline	1999-2014 (8 waves)	2002-2008	1983-2004
Age at baseline	20-85	43-74	35-64
Initial sample	43,611	34,876	unknown
Exclusion criteria	Not applicable	Self-reported hospitalization/diagnosis with AMI, stroke, coronary heart disease or angina. Or positive score on the Rose Angina Questionnaire	Documented or self-reported history of myocardial infarction or unstable angina pectoris
Analytic sample	43,611	23,511	97,048
Incident/ Prevalent CHD cases	Prevalent	Incident	Incident
Case ascertainment	Self-*reported "Has a doctor ever said you had a heart attack or CHD?"*	National MI and mortality registries	National MI and mortality registries
	Sensitivity analyses further restricted to "heart attack" only		
Cases (n)	1,933	309 (fatal) + 323 (non-fatal) = 632 total	6 522
Follow-up (median)	-	6.9 years	10.0 years
Statistical model	Logistic Regression	Cox Proportional Hazards Regression	Cox PH regression within each country & gender, followed by meta-analysis.
Weighting/ adjustment	Adjusted for age, sex.	Adjusted for age, sex, country	Adjusted for age
	Weighted to account for non-random sampling, response bias and clustering		

Table 5. Description of observational studies.

SD, standard deviation. CHD, coronary heart disease. MI, myocardial infarction.

2.2. Mediation and international differences

2.2.1. Rationale of the HAPIEE study, and this analysis

The Health, Alcohol and Psychosocial factors In Eastern Europe (HAPIEE) study is a multi-centre prospective cohort study of urban populations living in Russia, Poland and the Czech Republic. The rationale for study was described in a cohort description published in 2006,(245) as well as in the study's funding proposal (Box 1).

Briefly, the study was designed to investigate the role of nutritional, psychosocial and alcohol-related factors in the incidence of common disease (both at the within-country level, as well as to account for international differences in rates of disease). Previous publications from this study have reported mortality associations for nutrition,(161) alcohol consumption,(172) and socioeconomic status.(2) Nutrition and alcohol did not explain differences in rates of disease between cohorts. The current publication therefore assesses, in a systematic fashion, several original objectives of the study. I do not use data from Lithuanian arm of the HAPIEE study, as its baseline questionnaire did not include all the psychosocial variables of interest.

DETERMINANTS OF CARDIOVASCULAR DISEASES IN EASTERN EUROPE: A MULTICENTRE COHORT STUDY

Summary

The aim of the proposed research is to investigate the causes of the high rates of cardiovascular and other non-communicable diseases in the restructuring countries of central and eastern Europe. Specifically, we will test the hypotheses that cardiovascular mortality is related to (i) low dietary intake of fruits and vegetables and its biomarkers; (ii) high plasma levels of homocysteine and low levels of folate, possibly interacting with a polymorphism in MTHFR gene of folate metabolism; (iii) heavy alcohol consumption and/or binge drinking pattern; and (iv) unfavourable socioeconomic and psychosocial environment. This proposal builds on a successful validation study funded by the Trust, and on our previous work in the region. We propose to set up a multi-centre cohort study in Krakow (Poland), Novosibirsk (Russia), and 4 cities of the Czech Republic, with 30,000 men and women aged 45-69 randomly selected from population registers. At the baseline, exposures to suspected risk factors will be measured by questionnaires, short examination and analysis of blood samples. Participants will be prospectively followed up for mortality, using the national death registers, and for non-fatal events by annual postal questionnaires. In addition to the primary scientific purpose, the study will have an important training function by contributing to establishing a local research expertise.

Lay summary

Life expectancy in the restructuring countries of central and eastern Europe and the former Soviet Union is substantially lower than in the west. Most of the gap is due to high mortality from cardiovascular diseases and injuries. The aim of the proposed research is to study the causes of cardiovascular and other non-communicable diseases in eastern Europe. In particular, our hypotheses are that low dietary intake of fresh fruits and vegetables (and related substances in blood), high alcohol intake and binge drinking, and unfavourable socio-economic and psychosocial conditions are related to the high levels of cardiovascular diseases. On the basis of a successful pilot study, funded by the Trust, and our previous collaborations in the region, we propose to establish a prospective study of 30,000 men and women aged 45-69 years, recruited in Krakow (Poland), Novosibirsk (Russia), and 4 cities of the Czech Republic. At the baseline, we will measure their exposure to the suspected risk factors, and we will then follow these subjects for mortality. The study will be large enough to detect biologically meaningful effects of the risk factors. In addition to improving the understanding of the mortality crisis in eastern Europe and of the causes of cardiovascular diseases in general, the research will have an important training function, contributing to establishing a local research expertise.

i. AIMS OF THE PROJECT

In 1998, we applied to the Wellcome Trust for funding to establish a multi-centre cohort study in three countries of Central and Eastern Europe (CEE) and the former Soviet Union (FSU) to examine the causes of the high rates of cardiovascular disease (CVD) and other non-communicable diseases (NCD) in the region. The Trust recognised the scientific importance of the subject but suggested that we first conduct a pilot study to test the feasibility of the project and validate the measurements. The pilot study, summarised in the next section, was extremely useful; it established the feasibility of the study, answered the questions about the methodology, and provided solution for logistics problems.

The specific objectives of the proposed cohort study are to test the following hypotheses:

1) that psychosocial factors, both individual based and population based, are related to CVD and other non-communicable diseases;

2) that low consumption of fresh fruits and vegetables, and their nutrient biomarkers, are associated with increased risk of CVD;

3) that elevated concentration of homocysteine and low levels of folate and related B vitamins are related to CVD;

4) that the binge drinking pattern of alcohol consumption is related to CVD (and injury);

5) that CVD risk is affected by interaction between different groups of risk factors, in particular between heavy drinking and folate deficiency, and between the MTHFR genotype (of folate metabolism) and folate deficiency.

The high rates of NCD, large variation in exposure, and low cost of data collection provide an ideal opportunity. In the five years covered by the proposed programme support, the cohorts would be established, and data from the cross-sectional phase and early longitudinal stage (the first 3 years of follow up) will be analysed. Further follow up, enabling the more definite test of aetiological hypotheses, will continue beyond the 5 years covered by the proposal, using administrative records and tracing death certificates. If the success of the pilot study is translated into a success of the main study, it would form a basis of a centre for training in non-communicable diseases in the restructuring countries.

Box 1. Uninterrupted extract of the first two pages of the original funding proposal. *This was submitted to the Wellcome Trust in 2001. Key sentences relevant to this thesis are highlighted in darker font.*

2.2.2. Participants

Random population samples of 28 945 men and women aged 43–72 years at baseline in 2002–2008 were selected from population registers and electoral lists. The overall response rate was 59% (61% in Russia and Poland; and 55% in the Czech Republic). I excluded 7173 participants (25%) with a history of Acute Myocardial Infarction, Stroke, Chronic Heart Disease or Angina, as well as those who scored positively on the Rose Angina Questionnaire. A further 905 participants (4.2%) were lost to follow-up, primarily since they did not give consent to link their records with the national mortality registries, or due to migration (losses of 4%, 7% and 1% in Czech, Polish and Russian samples, respectively). This left an analytical sample of 20 867 participants. The study was approved by the University College London/University College London Hospital ethics committee and by the local ethics committee in each participating centre. All participants gave written informed consent.

2.2.3. Cardiovascular outcomes

In the Czech Republic, linkage was done with the national mortality register, in Poland with the regional (Voivodship) mortality register, and in Russia with Novosibirsk city mortality register. Participants were linked to deaths records using a national personal ID number in the Czech Republic and Poland. In Russia, linkage was done using surname, initials and date of birth, where any potential inconsistencies were corrected manually. 905 participants (4.2%) were lost to follow-up, primarily since they did not give consent to link their records with the national mortality registries, or due to migration. Political developments and subsequent changes to Russian law meant that the study team (both Russian and international colleagues) were unable to access mortality data after 2011. For this reason, Russian follow up is slightly shorter. Follow-up was available to 31st December 2010 in Russia and Poland, and 30th June 2013 in the Czech Republic (maximum follow up of 8.0, 8.9, and 11.3 years respectively). The primary end-point was CVD mortality (ICD-10 codes I00-99).

2.2.4. Socioeconomic factors

At baseline, participants completed an extensive structured questionnaire, and underwent a standardized nurse examination in a clinic. English versions of the exact questionnaire items used are shown in Annex 1, with a brief summary below. *Highest educational qualification* was categorized as primary or less; secondary and tertiary. Assuming that the average participants in these groups were separated by around three and six years of additional schooling respectively, I modelled a linear relationship between these three categories. For self-reported *economic activity*, participants were classified into three groups: economically active; retired and no longer working; and currently unemployed. Participants were also asked about long-term unemployment, with 4 options (never; up to 3 months total, up to 3-12 months total, more than 12 months total); responses were dichotomized, comparing the extreme group to the rest.

Possession of 10 *material amenities* (microwave, video recorder, colour television, washing machine, dishwasher, freezer, camcorder, satellite TV, telephone and mobile phone; each coded 0/1) was used to derive a standardized continuous scale. For *current material deprivation,* three questions asked about how often the subjects had difficulties with paying for food, clothes, or bills. Each item was scored 0-3 to yield a total score of 0-12. This was modelled per one standard deviation greater material deprivation. All standardization procedures used the mean and SD obtained from all three cohorts in combination (as opposed to standardizing, so that the standardized mean in each country is zero). Similarly to *current material deprivation*, another scale of *early-life material deprivation* asked about the same items, this time with participants retrospectively recalling their childhood.

In addition, participants were asked whether the *"Changes since 1989 have been good or bad for your general social position?"* Original responses in 5 categories (Very good, good, no change, bad, very bad) were regrouped into three categories.

2.2.5. Psychosocial factors

Social support is typically measured with a multifaceted instrument, covering aspects of friends, family, social clubs and marital status. Given the larger power of this study, I did not combine these separate domains. For *marital status*, I compared the "married/cohabiting" reference group against "divorced/widowed", or "single". "Are you a member of a *club*/organization" was kept binary. *"How often do you see relatives outside of your household?"* originally had 6 options (*"several times a week"; "About once a week"; "Several times a month"; "About once a month"; "Less than once a month"; "I don't have any relatives"*). I dichotomized this at the "at least monthly" threshold. A separate question *"How often do you see friends outside of your household?"* was handled similarly.

Depressive symptoms were assessed with the CESD-20 questionnaire (range 0-60). A binary trait for depressive symptoms was defined as CESD-20 score ≥ 16. *Perceived control* was assessed with a scale developed by the MacArthur programme on midlife development.(247) The subscale "control over life" (range 0-40) was used as a standardized continuous variable.

2.2.6. Conventional CVD risk factors

Covariates measured at baseline included country, age, gender, and eleven conventional CVD risk factors: smoking status (5 categories of never smoker; ex-smoker; 1–10, 11–20, and 21+ cigarettes a day), diabetes, and physical activity (dichotomized at 2.5 hours a week) were determined by self-report. Clinical examination[13] determined systolic blood pressure (modelled linearly from 115mmHg onwards)(248); as well as BMI, total and HDL serum cholesterol (whereby all three were modelled with linear and square terms after centring at 23 kg/m^2, 6 mmol/L and 1.5 mmol/L, respectively). Alcohol intake in the last 12 months was self-reported using the graduated frequency questionnaire (GFQ).(249) The UK Chief Medical Officers' alcohol guideline advises men and women to consume less than 14 units of alcohol (equivalent to 112 grams of ethanol) per week. Participants were categorized as either non-drinkers, drinking within these guidelines, drinking up to twice the guideline limit, or drinking more than twice the guideline limit. For completeness, I included three additional alcohol-related covariates. Frequency of alcohol consumption was dichotomized at the once a week threshold. A pattern of binge drinking was defined if men/women reported consuming more than 100g/60g of ethanol in one episode at least monthly. The CAGE questionnaire was used to evaluate symptoms of problems with alcohol, and was dichotomized at 2 or above.

2.2.7. Statistical analyses

2.2.6.1. Missing data

Between 0-12% of the data was missing for each variable (Table 10). This was imputed from 10 multiple imputation models that included vital status, follow-up time, all covariates. The predictor matrix additionally included key variables that were associated with missing variables and their missingness (from either the baseline survey, or a follow-up survey done 3 years later on the same participants). These predictors were further self-reported details about the time that participants spent on sports, games and hiking; the tendency to rely only on self (as opposed to others); over-commitment at work; amount of trust in the local area; perceptions that money influences health; perceptions that others treat the participant unfairly; unemployment of others in the participant's household; hypercholesterolemia; antihypertensive drug and statin usage; and subsequent depression in wave 2.

2.2.6.2. Main analysis

Cox regression was used to estimate hazard ratios (HR) and 95% confidence intervals (CIs) for associations between psychosocial factors and mortality end-points, using follow-up time as the time scale. Data were pooled across three cohorts and both genders. Three models were created with increasing levels of adjustment. Model 1 was adjusted for age, sex and country. International differences in cardiovascular mortality in this region are known to be much larger in men than women. To examine potential causes of this, I looked whether gender and/or being placed in Russia modifies the association between psychosocial risk factors and CVD mortality. I included interactions that met a Bonferroni-adjusted P-value threshold of 0.05/14 = 0.0036. In model 2, I additionally adjusted for eleven conventional CVD risk factors for two reasons. First, as an indication of how much of the hazard from psychosocial exposures might plausibly be mediated indirectly via conventional risk factors, and, second, as an indication of their direct effect size, through pathways not measured in this study. Model 3 was additionally adjusted for the six

psychosocial factors that associated with CVD mortality (at $p<0.05$) following backwards-stepwise elimination from a larger model that began with all 14 candidate psychosocial factors. The proportional hazards assumption was tested using Schoenfeld residuals and checked graphically with log-log plots, with no evidence of its violation.

In all three models, I assumed no differences in the size of the risk factor-mortality associations between participants located in Poland, the Czech Republic and Russia. There was insufficient power to explore heterogeneity between Poland and the Czech Republic, but I examined effect modification between participants from Russian vs. Czech/Polish cohorts. Additional sensitivity analyses looked for effect modification by gender; repeating the main analysis after excluding imputed data; excluding participants with less than 2 years of follow-up (to reduce reverse causation bias); and using a similar 8-year follow-up for all three countries.

2.2.6.2. Attenuation analysis

Simpler and complex models were compared with each other, to evaluate the degree of attenuation, and thereby infer approximate degree of mediation, using the formula:

Attenuation/amount mediated = [log(Fully adjusted HR) - log(Crude HR)] / [log(Crude HR)].

For example, if the association between education and mortality is HR=1.45 in model 1, and this attenuates to HR=1.20 in model 2, then one can infer that the conventional cardiovascular risk factors (which are additionally included in model 2 but not in model 1) might account and potentially mediate around half of the pathway from education to mortality. Of note, this is a relatively crude method which is prone to differential measurement error as well as model misspecification. Accordingly, I used this method only to provide very approximate and qualitative inferences

about whether putative mediators are likely to play a small or large role, respectively.

When investigating the causes of international variation, literature suggested that these may be larger in men than women. Accordingly, I set up models where the reference group were women situated in either Poland or the Czech Republic. Mortality rates here were compared against three other groups: men from Poland or the Czech Republic (i.e. the pure gender effect); women from Russia (i.e. the pure country effect), and men from Russia (i.e. the interaction of gender and country combined).

2.2.6.2. Population Attributable Fraction

The Population Attributable Fraction denotes the proportion of mortality which could be prevented, if the entire population were not exposed to a given risk factor (and assuming a causal relationship between exposure and outcome).(250) In my study, it was calculated by fitting model 2 to the first set of imputed data, and then using *punafcc* command in STATA 14. As this package is unable to handle continuous risk factors, the continuous variable *material amenities* was dichotomized into two halves using the median as a cut off. For education, I calculated attribution if everybody without tertiary education would attain tertiary education. BMI was dichotomized as obese (BMI>30kg/m^2) or not. The STROBE checklist was followed for presentation.

2.3. Prediction

2.3.1. Description of the derivation dataset (HAPIEE)

2.3.1.1 Sample and baseline data

The HAPIEE study data collection methods were described in section 2.2.1. In this section of my thesis, I excluded participants from Lithuania due to later entry into the study, slight differences in data collection, and shorter duration of follow-up. I excluded participants who self-reported a diagnosis or previous hospitalization for angina, myocardial infarction or stroke, as well as those who scored positive on the Rose Angina questionnaire (altogether 25%). Participants were included if they had chronic kidney disease, peripheral artery disease or diabetes, since CVD prevention is may be suboptimal for such participants in this region. I further excluded 9% of participants taking cholesterol lowering medication (mainly statins) at baseline, as they were unlikely to benefit from risk prediction and stratification. 23% of participants had missing data on at least one of variable. Between 9% to 13% of the observations were missing for depression, blood pressure, cholesterol and BMI. For other variables, missing observations formed less than 1.1% of each variable. Altogether, 23% of persons with some missing data were removed, so the analytical sample was a complete case sample.

2.3.1.2 Outcome data

International variability in coding nonfatal outcomes is likely to reduce data quality. For this reason, I adopted the same method as the original SCORE model, and only considered fatal outcomes. There is also international variability in the coding of smaller causes of death. Hence I took the primary outcome as *any CVD mortality* (ICD-10 classification: I00 to I99).

Median follow-up was 6.6, 7.1 and 9.6 years for Russia, Poland and Czech Republic. Time after 10 years of follow-up was censored. There were 102 female CVD deaths, and 236 male CVD deaths in the final analytical derivation sample.

2.3.2. Description of the external validation dataset (Estonian Biobank)

2.3.2.1 Sample

External validation was performed using data from the Estonian Biobank, whose details have been published previously.(251) Briefly, baseline data were collected between 2002-2011 (median date of recruitment = 04/11/2008), from 51 141 participants aged 18-103 (median age = 43), typically by nurse practitioners at the patients' primary care centre. The sampling frame can be thought of as a hybrid between a conventional population-based approach and a volunteer-initiated approach. On the one hand, no invitation letters were sent to participants. Instead, participants heard about the study either at special promotion events, from the media, from friends, or at their healthcare provider. As the sample was not selected from the general population, this does not enable the calculation of a response rate. It is plausible that when compared to the general population, the Biobank sample was enriched for healthy participants with higher SES (just as is seen in most population-based cohort studies). However, it is also plausible that recruitment via healthcare providers would have additionally enriched the sample from the other extreme: those with symptoms and early disease. Hence it is possible for the net all-cause mortality rate of the sample to be closer to the population average, when compared to conventional population-based cohort studies.

Another dimension which increases the generalizability of the study is the wide coverage of recruitment sites. 56% of all the registered primary care physicians were involved with recruitment, spanning each of the 15 administrative countries within Estonia, (figure 24). Hence the sample involves a mixture of urban and rural participants.

Figure 24. Map of Estonia's 15 administrative counties, illustrating the recruitment centres of the Estonian biobank study. *From (251).*

In comparison to the derivation cohort, the validation cohort has a different sampling frame (urban vs. urban-rural, respectively); different invitation method (postal invite only vs. opportunistic recruitment and self-referrals); and slightly different data sources to ascertain past medical history and current medication usage (discussed below).

2.3.2.2 Data collection

The derivation and validation cohorts both collected the majority of data using similar methods: a trained nurse performed computer-assisted personal interviewing, physical examination and blood tests. Past medical history and medication use was assessed by self-reported questionnaire in the derivation cohort (in this context, to determine baseline diabetes, CVD, antihypertensive and statin use). In the validation cohort, most nurses additionally had access to the participant's primary care records.

Furthermore, the study team also performed linkage to the national acute myocardial infarction registry, as well as to the national electronic healthcare and prescriptions database. As of Dec 2011, approximately 80% of inpatient discharge letters and 80% of all dispensed prescriptions were recorded in this system.

Plasma cholesterol was measured by Proton NMR spectroscopy (Brainshake Ltd.) in Finland, on a random subset of the full cohort.(252) Cholesterol measures were available for 22% of the participants, meaning that cholesterols were missing completely at random for 78% of participants. Other variables were missing (at random) for up to 24% of the participants. For physical inactivity, employment and marital status, data were missing for 19% 6% and 4% of participants, respectively. All other variables were at least 99% complete. I performed the validation on the complete case sample. This discarded the majority of participants who had missing cholesterol. However, since the cholesterol was a completely random subsample, it is almost impossible for this decision (to omit participants with missing cholesterol) to lead to sampling bias.

After median 7.6 years of follow-up, 91 CVD deaths were detected by linkage to the national mortality registry among the 4632 participants in the validation sample.

2.3.2.3 Age selection

Since Estonia is considering the development of a national CVD screening programme, I validated the newly-derived models as close as possible to eventual real-life settings where these may be used. One key variable is the age range among whom to conduct screening. In countries like the UK, screening for CVD among asymptomatic individuals is typically done for men aged >40 and women aged >50. Similar recommendations have been made by the European Society of Cardiology.(253) Given higher baseline rates in countries like Estonia, then recalibrating these Western European thresholds into the Estonian context equates to an equivalent

absolute level of risk as seen in men aged >37 and women aged >47 in Estonia (Figure 26), which I took as the lower age limits in my validation study. For the upper age limit, it is unclear at what age any geriatric risks associated with statins overshadow the cardiac benefits. In the meantime, a common upper threshold in other models is 74 years, which I also applied to Estonia.

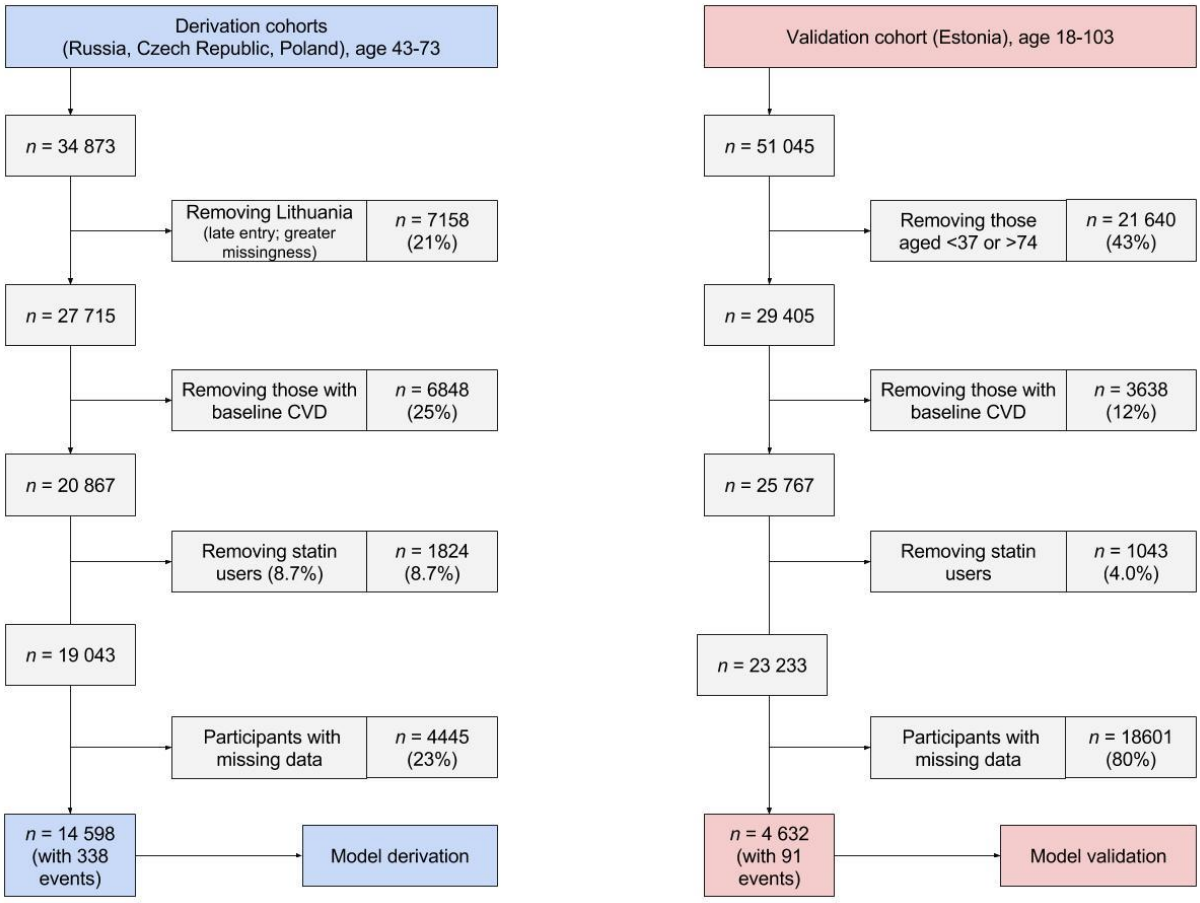
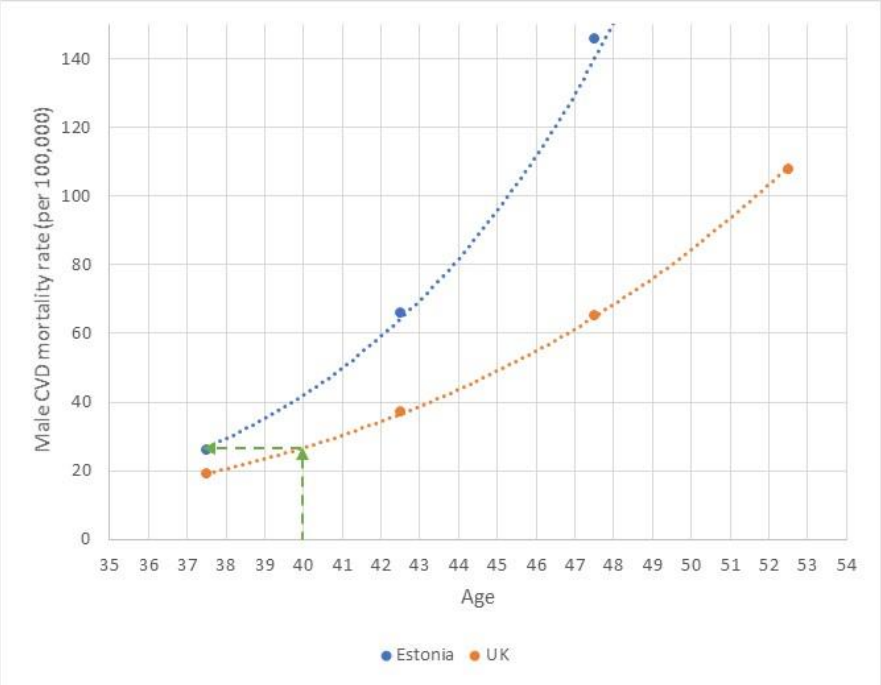


Figure 25. Flowchart illustrating sample selection in the derivation and validation cohorts.

		Male CVD mortality	
Age band	Mean age	Estonia	UK
35-39	37.5	26	19
40-44	42.5	66	***37***
45-49	47.5	146	***65***
50-54	52.5	271	***108***

		Female CVD mortality	
Age band	Mean age	Estonia	UK
35-39	37.5	7	8
40-44	42.5	14	14
45-49	47.5	30	23
50-54	52.5	60	***38***



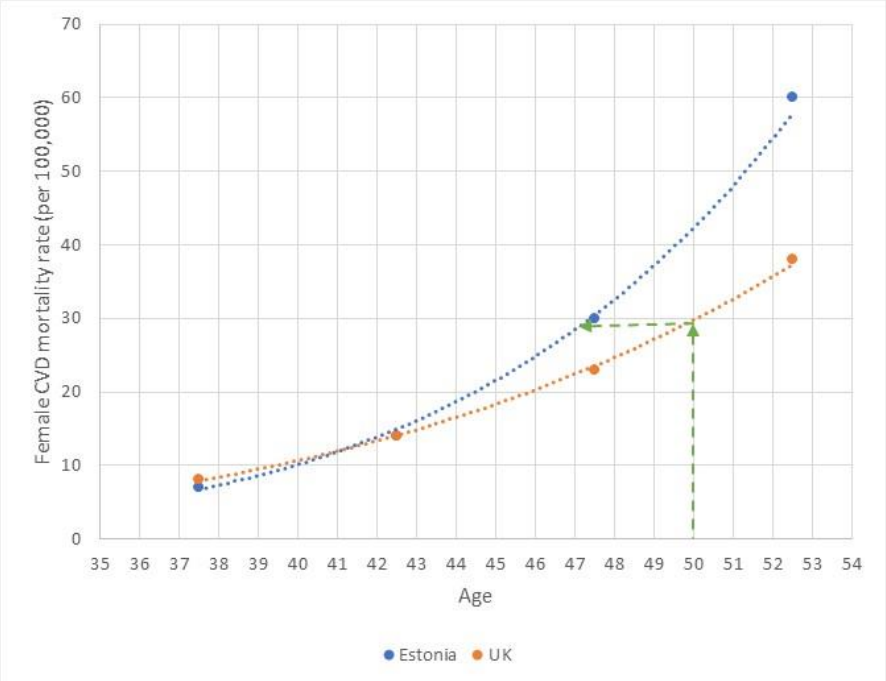


Figure 26. Rates of CVD mortality, per 100 000 inhabitants, in Estonia and the UK. *Data are from the Global Burden of Disease study, which relies predominantly on WHO Mortality data.(178) Bold italic font denotes age bands where screening for CVD has been suggested/attempted among Western European countries.*

2.3.3. Predictor harmonization and selection

2.3.3.1. Data harmonization

Data were coded similarly across the two datasets for the following variables: age, sex, diabetes, systolic blood pressure, total cholesterol, Body Mass Index, employment and marital status. Of note, age was truncated to the nearest whole year in Estonia for privacy reasons, potentially leading to greater measurement error in the assessment of chronological age.

To categorize smoking status, I combined data from three self-reported questions: self-reported smoking status, current smoking intensity, and smoking history as follows:

Category	**Definition**
1) Never-smoker	Self-reported never smoker
2) Ex/light smoker	Ex-smoker OR Current smoker who smokes <1 cigarette per day OR Current smoker who has smoked < 365 cigarettes to date
3) Current smoker	Smokes ≥ 1 cigarette per day AND Has a smoked ≥ 365 cigarettes to date

Educational qualifications were categorized as either “primary” (typically leaving school at the age of 16), “secondary” (typically leaving school at

the age of 18), or “tertiary” (a university or higher vocational degree), after consideration of the ISCED97 criteria in each of the 5 countries.

In the validation dataset, depression was assessed by a single-item screening question, completed by self-report in the baseline questionnaire, with three options: “I do not have anxiety/depression” (reference), Vs. “I have moderate anxiety/depression” or “I have severe anxiety/depression”. The final two options were merged in my analysis to derive a dichotomous “possible depression” variable. In the derivation dataset, an equivalent “possible depression” variable was created by dichotomizing scores from the more comprehensive CESD-20 inventory.

I initially sought to also harmonize alcohol consumption, as this is a plausibly novel risk factor for the development of heart disease. However, as the validation sample was much smaller (after discarding participants with missing data on cholesterol), there were insufficient CVD events among those participants who reported consuming more alcohol than advised. Furthermore, previous publications from the derivation sample have highlighted how various indices of alcohol excess do not associate strongly with subsequent CVD.(172) For this reason, alcohol was not considered.

2.3.2.2. A priori predictor selection

After harmonizing all risk factors, the list of candidate predictors was selected as follows. I selected all risk factors used in other CVD risk prediction models. I added those risk factors that have meta-analytic evidence of independent associations with cardiovascular disease. I eliminated invasive markers (e.g. those requiring blood tests). This resulted in seven candidate predictors which are not present in the conventional SCORE model. Four of these new predictors were psychosocial/socioeconomic (education, marital status, employment status, depression) and three were biological or behavioural (BMI, physical inactivity and antihypertensive use).

2.3.4. Derivation of new models

2.3.4.1 Data pooling

The ratio of events-to-parameters was low enough in some countries, that this risked causing overfitted models, if I fitted separate models in each country. As a solution, I divided the region into two halves. Cardiovascular mortality rates are higher in those Eastern European countries that were previously members of the Soviet Union (Estonia, Latvia, Lithuania, Belarus, Ukraine, Georgia) as well as those that border the Black Sea (Romania, Bulgaria). I allocated participants from these countries, with a higher baseline risk, an additional dummy variable ("high risk region"), which was compared against the reference dummy ("low risk region") for countries from predominantly Central Europe that border Germany and Austria. My models assumed that individuals from each of these two regions are otherwise indistinguishable (that is, I did not fit random effects or country-level dummy variables). In relation to the particular data at hand, this meant that I grouped participants from Poland and Czech Republic into a similar baseline risk category. This assumption appears reasonable, given the nearly parallel trends in WHO mortality for these two countries over the past 25 years (figure 27).

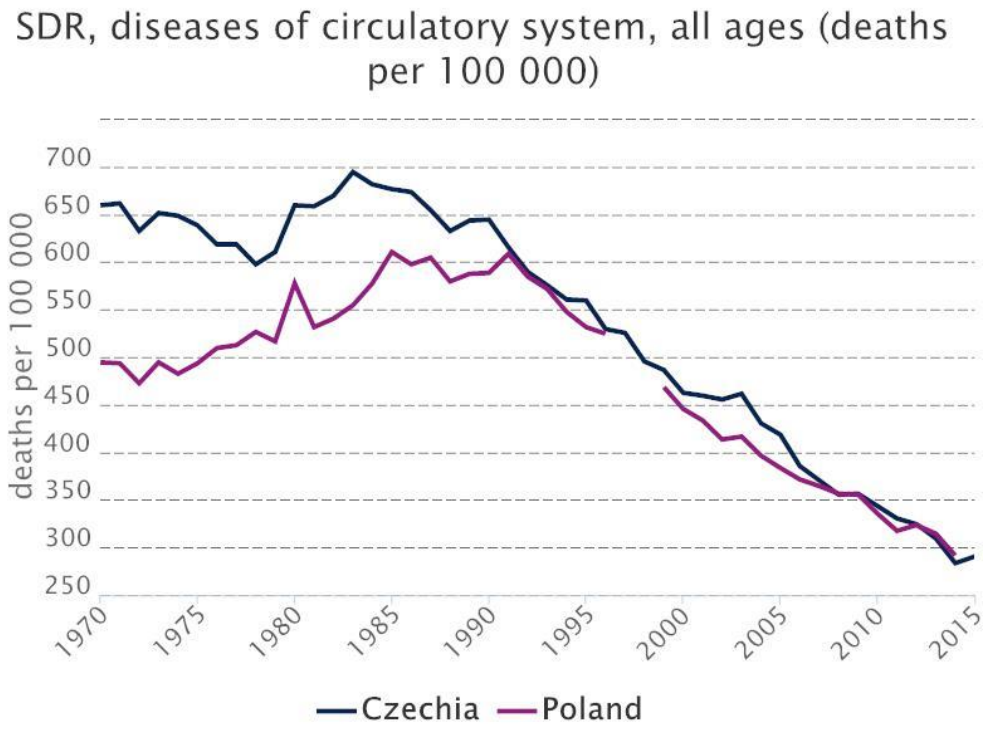


Figure 27. WHO mortality rates for cardiovascular disease in Poland and the Czech Republic.

Source: https://gateway.euro.who.int/en/hfa-explorer/#9aZl40c8fU

Data: WHO Health For All Database

The original SCORE model created separate models for men and women. In statistical terms, this is equivalent to fitting gender-by-risk factor interactions for all risk factors. Similarly, fitting separate models for high/low risk countries is equivalent to fitting area level interactions for all risk factors. While such approaches can improve risk prediction, this requires sufficient events in each arm to avoid overfitting the model. In my smallest arm for model development, I had just 43 female events in the high-risk region (Russia), making such a standalone model inappropriately small for my aims. Instead, I reasoned that the added benefit from gender-by-risk factor interactions, and area-by-risk factor interactions, would be relatively small. I.e. I assumed that smoking increases the Relative Risk of CVD for women just as much as it does for men, and that this magnitude also should be similar across low- and high-risk regions alike. However, since international differences CVD tend to be larger in men than women,

I hypothesized that international differences in the baseline rate (among those with no CVD risk factors) should be larger in men than women. Accordingly, I fitted a single gender-by-area interaction term (male*high-risk-area). Overall, my model development strategy was one where I made certain assumptions and brought associated limitations (e.g. I assumed homogeneous effects between risk factors and outcome across genders and regions), but this allowed me to derive my models on larger samples where the risk of overfitting would be lower.

2.3.4.2. Model 1 ("ORIGINAL-SCORE")

I used the high-risk SCORE model,(185) where two parametric Weibull models generated the baseline hazard function for men and women, respectively. Here, the risk of CVD increased with the presence of one of three risk factors. Risks from cholesterol increased linearly from 6 mmol/L upwards (with no difference in risk among those whose cholesterol was less than 6 mmol/L). Similarly, risks for systolic blood pressure increased linearly from 120 mmHg. Smoking status was dichotomized.

2.3.4.2. Model 2 ("EastEur-SCORE")

Cox regression was used to estimate new baseline hazards (using the R command *survest*), and coefficients for each of the SCORE risk factors (using the R command *coxph*), in a pooled dataset with data from both genders and both areas. My literature review suggested that additional benefit may be gained from more sophisticated modelling of the conventional SCORE predictors. Accordingly, I categorized smoking status into three as opposed to two categories (detailed in section 2.3.3.1).

I also tested quadratic functions to represent non-linear associations between cholesterol and CVD risk, and age and CVD risk. An *a priori* criterion for keeping these were that the quadratic should improve the LR test P-value by <0.05 and Harrell's C-statistic by >0.0001. Age-squared

had no benefit and was dropped, but cholesterol-squared was kept after centring at 6 mmol/L. An LR test comparing models with and without cholesterol-squared gave $p<0.001$, and $\Delta C=0.003$). Since cholesterol-squared has not been previously used in risk prediction settings, I further checked the plausibility of this non-linear association, by plotting a detailed spline plot in the derivation data (figure 28). A similar curved relationship was also visible in the validation data (data not shown).

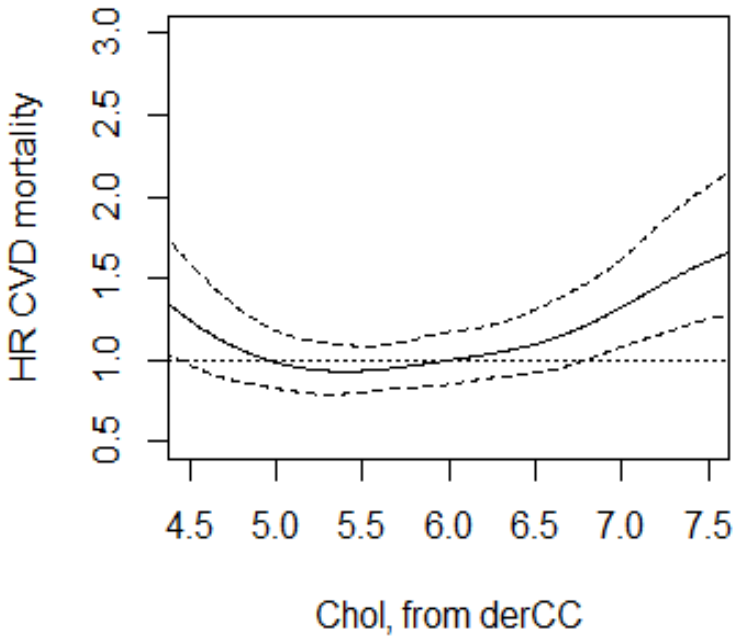


Figure 28. Association between total serum cholesterol (mmol/L, x-axis) and Hazard Ratio (HR) of CVD mortality (y-axis) in the HAPIEE derivation sample. *Data are from a cubic spline model, restricted to 4 knots.*

The spline plot suggested that the lowest risk of CVD may be found in those whose cholesterol is between 5.0 to 6.5 mmol/L. Since the original SCORE specification (6.0 mmol/L) falls within this range, I am happy to continue using a squared term that is centred at 6.0 mmol/L.

I also considered using a risk function that is more complex than a simple quadratic. However, since splines are difficult to interpret and transfer via paper to external researchers, and also as the spine plot above almost approximated a quadratic shape, then I was happy to keep a relatively simple quadratic risk function. With 10 coefficients in the final EastEur-SCORE model, I had 34 events per each coefficient. This is much larger than the recommended 10 events per coefficient, meaning that there was a lower risk of overfitting the model.

2.3.4.2. Model 3 (“HAPIEE-SCORE”)

Seven candidate predictors were added to the newly created EastEur-SCORE model, to create the final HAPIEE-SCORE model. BMI was modelled with linear and square terms, as previously suggested(254). Physical inactivity was dichotomized at <150 minutes of activity, when combining recreational and occupational activity levels. Educational attainment was modelled linearly across three categories, where the hazard of going from tertiary to secondary education is equivalent to the hazard of going from secondary to primary-or-lower education. Employment and marital status were coded to three categories (employed, unemployed, retired) (single, widowed/divorced, married/cohabiting). Possible depression was dichotomized at those scoring 16 or above on the CESD-20 questionnaire. Finally, the use of antihypertensive medications was dichotomized, and following the QRISK model I added an interaction between antihypertensive use and systolic blood pressure.

I also explored the potential for interaction terms across the risk factors against three core variables: age, sex and education. For some of these interactions, the P value was 0.02. However, as these interactions arose from 21 multiple tests, I was concerned that these could have arisen by chance alone, and so decided to not include these. The final model had

21 coefficients. With 16 events per each coefficient, the possibility for overfit remained moderate.

2.3.5. Evaluating the performance of single models

Calibration plots were used to determine the degree of over/under-prediction among participants grouped into six risk strata. The size of each strata was determined by dual consideration of having at least 10 events per strata, as well as clinically meaningful thresholds.

Discrimination was assessed by Harrell's C-statistic, using the R package *rms* (specifically, the commands *cph & survest*). 95% confidence intervals were calculated manually, by extracting standard errors from the package *Hmisc* (specifically, the command *rcorr.cens*).

Classification ability was assessed by dichotomizing participants' risk at the 5% risk threshold for CVD mortality, a commonly suggested threshold above which to consider clinical interventions like statins.(253) This dichotomized approach allowed the calculation of sensitivity, specificity, PPV and NPV as per usual conventions.

Given how I defined the clinical threshold of interest as 5%, this was used to infer an appropriate weighting for the *Net Benefit* calculation. This was made with the initial assumption that benefit gained from correctly identifying one additional True Positives is approximately 20x as important, when compared to the cost of incorrectly labelling one additional person as a False Positive. This is based on the rationale as proposed by Vickers.(255) This assumption was relaxed by the visual use of Decision Curve Analysis, where the clinical intervention threshold was varied from 0% to 20% (denoting a Net Benefit weighting that ranges from infinity to 5x, respectively).

2.3.6. Evaluating change across two models

Change in calibration was assessed visually. Change in discrimination was assessed by *ΔC*. Change in classification was assessed by *ΔNet Benefit* and *Decision Curve Analysis*.

As sensitivity analyses, *Reclassification Plots* were used to inspect reclassification performance across the entire risk spectrum, and to gain a better insight about changes near and far away from clinical thresholds. Despite my reservations around using the Net Reclassification Improvement (NRI, detailed in Annex 2), given the popularity of this metric in the literature I will nonetheless present *continuous NRI* and *binary NRI* (also known as categorical NRI) stratified for cases and controls, using the R package *PredictABEL*.

2.3.7. Modelled clinical effect from statins

It was beyond the scope of this thesis to produce a formal cost-effectiveness analysis of these models, when scaled to real-life settings. Nonetheless, I wanted to briefly explore the viability of CVD screening for a country such as Estonia. The purpose of this was not to provide reliable estimates of anticipated consequences. Rather, this was done as a sensitivity analyses, which sought to identify how the risks and benefits of an intervention such as statins, when used across a range different clinical risk prediction models. It was beyond the scope of the thesis to formally test a range of parameters and assumptions built into the model. Instead, I calculated changes to costs, clinical outcomes and cost-effectiveness for just one hypothetical scenario with the following assumptions:

- It costs around €50 of clinical care time to initiate statin treatment among one person of high risk. This is followed with €2/month for the price of the statin itself. Altogether, this amounts to €170 per person over 5 years. I assume 100% adherence to treatment in my models.
- Statins lower LDL cholesterol by 2 mmol/L, which would lower CVD mortality by 23% (95% CI = 17 to 29%) (256) among the true positives over five years. One year of CVD mortality averted is equivalent to living with 100% quality of life.
- Statins additionally cause benefits by preventing the development of non-fatal disease. However, as some of these later transition into fatal events within a 5-year window, then this requires more sophisticated modelling. Consequently, I did not model nonfatal benefits.
- Statins cause serious adverse events (predominantly diabetes) in 90/10 000 people treated over 5-years. I assume that these consequences are immediate and lifelong (i.e. lasting throughout the 5-year window of my model), whereby one year of such illness is equivalent to 75% of full health.

Another way of illustrating this calculation is to consider a hypothetical dataset where 10 000 people are screened, following which 3000 are treated with statins. Of these 3000, 300 people are correctly treated since they would otherwise have developed the event, while the remaining 2700 are potentially over treated, since they do not develop the event in the follow-up period. Of the 300 who are correctly treated, this may lead to 69 events averted (since 300 * an intervention effect of -23% = 69 events averted). From this:

The ratio of (Correct treatment : total treatment) [CT:TT] = 300÷69 = 4.34

2.3.7.1. Number Needed to Treat

Following the example above, the Number Needed to Treat (NNT) denotes the number of people required to treat, to prevent one CVD fatality.

*NNT = (N treated * CT:TT) ÷ (N correctly treated)*

*NNT = (3000 * 4.34) ÷ 300 = 43*

Optionally, this equation can also be rearranged for simplicity as:

NNT = N treated ÷ (N of cases among those of predicted high risk × 0.23)

2.3.7.1. Number Needed to Screen

Following the example above, the Number Needed to Screen (NNS) denotes the number of people required to screen, to prevent one CVD fatality.

*NNS = (N screened * NNT) ÷ (N identified as high risk)*

*NNS = (10000 * 43) ÷ 3000 = 145*

Optionally, this equation can also be rearranged for simplicity as:

NNS = N screened ÷ (N of cases among those of predicted high risk × 0.23)

A summary of the relationships is illustrated in table 6.

	Hypothetical cohort	Calculations	
Nr of events averted	69	1	
Nr treated, correctly	300	4.3	= CT:TT
Nr treated, correctly + incorrectly	3000	43	= NNT
Nr screened	10000	145	= NNS

Table 6. Hypothetical illustration of the calculation of the ratio of Correct Treatment to Total Treatment (CT:TT), the Number Needed to Treat (NNT), and the Number needed to Screen (NNS).

3. Results

3.1. Mendelian randomization

3.1.1. Conventional Mendelian randomization

Using conventional MR analysis, 1-SD longer education (due to genetic predisposition across 162 SNPs) was associated with a 33% lower risk of CHD (OR= 0.67 [0.59 to 0.77]). Figure 29 additionally shows individual causal estimates from each of the 162 SNPs.

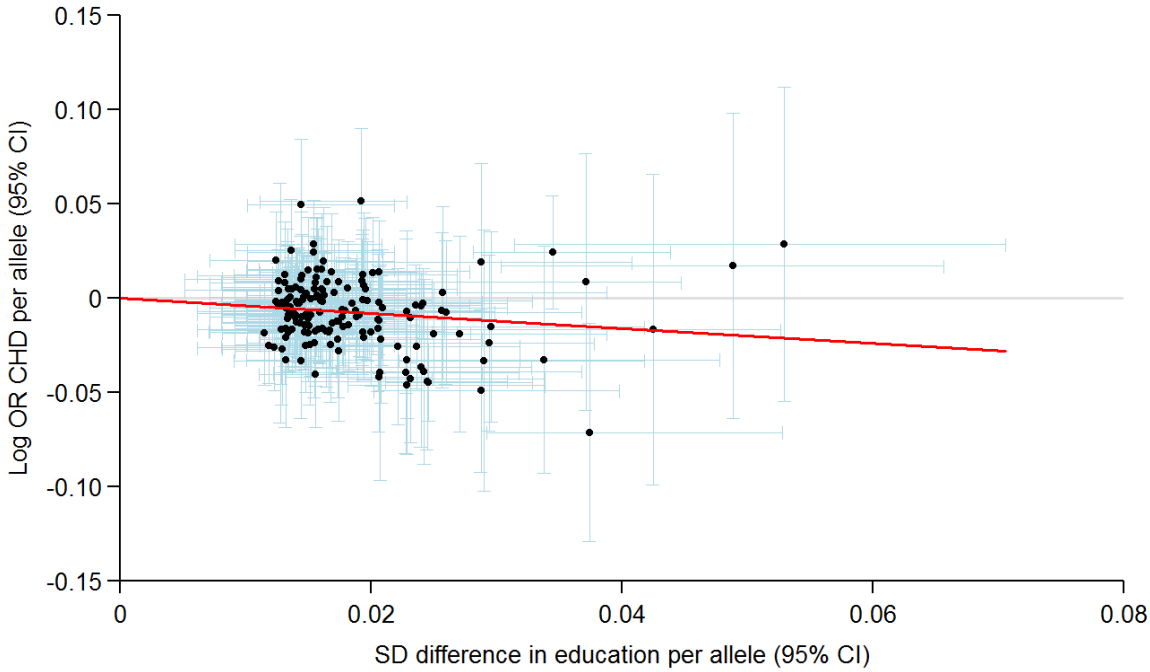


Figure 29. Scatter plot of 162 SNPs associated with education and their risk of CHD (with 95% confidence intervals).
Each dot represents one single nucleotide polymorphism (SNP).
The red line represents the regression slope of the causal effect estimate of education on risk of CHD (where each SNP is weighted by its inverse allele frequency). CHD, coronary heart disease. OR, odds ratio. CI, confidence interval.

A secondary set of analyses using a set of 72 SNPs instead of 162 SNPs yielded consistent results in terms of direction, magnitude and statistical significance (OR=0.60 [0.49 to 0.74]). Results from each SNP are shown in figure 30.

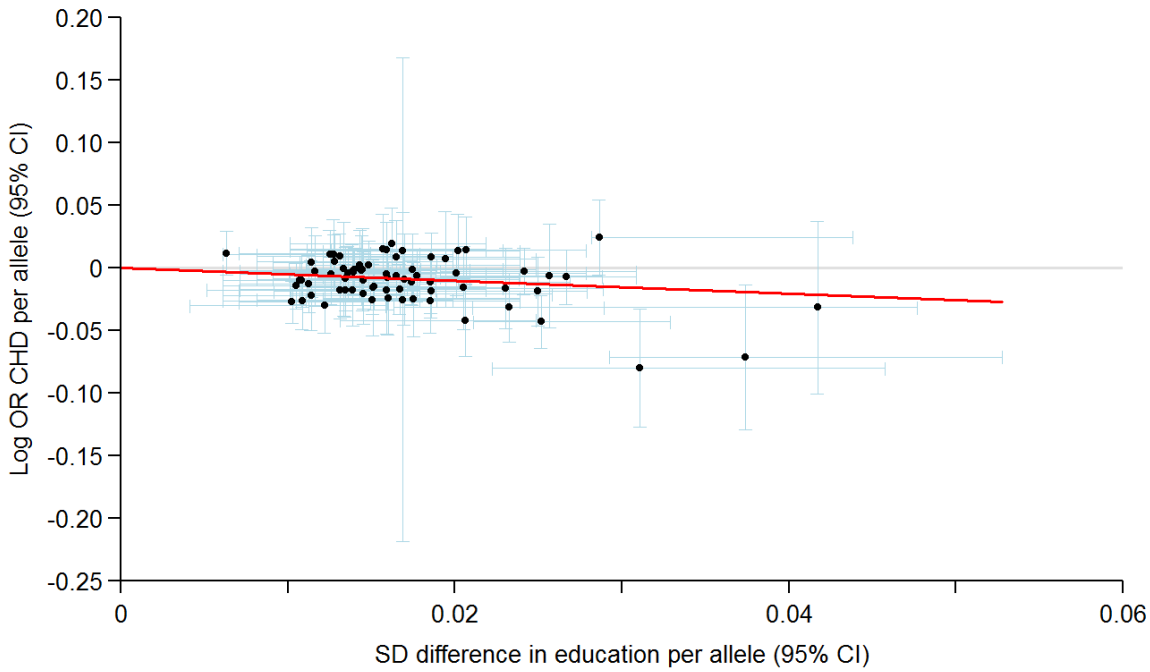


Figure 30. Scatter plot of the 72 SNPs associated with education and their risk of CHD (with 95% confidence intervals).
Each dot represents one single nucleotide polymorphism (SNP).
The red line represents the regression slope of the causal effect estimate of education on risk of CHD (where each SNP is weighted by its inverse allele frequency). CHD, coronary heart disease. OR, odds ratio. CI, confidence interval.

3.1.2. Sensitivity analyses for pleiotropy

As expected, sensitivity analyses using MR-Egger and weighted median MR provided less precise estimates than with conventional MR. Nonetheless, their causal estimates were similar in terms of direction and magnitude, and were unlikely to have happened by chance alone (Figure 31). There was little evidence of a non-zero intercept from the MR-Egger test (intercept beta = 0.004 [-0.056 to 0.013]; P-value=0.417), consistent with the hypothesis that pleiotropy was not driving the result. The MR-regression slopes are illustrated in Figures 32 & 33.

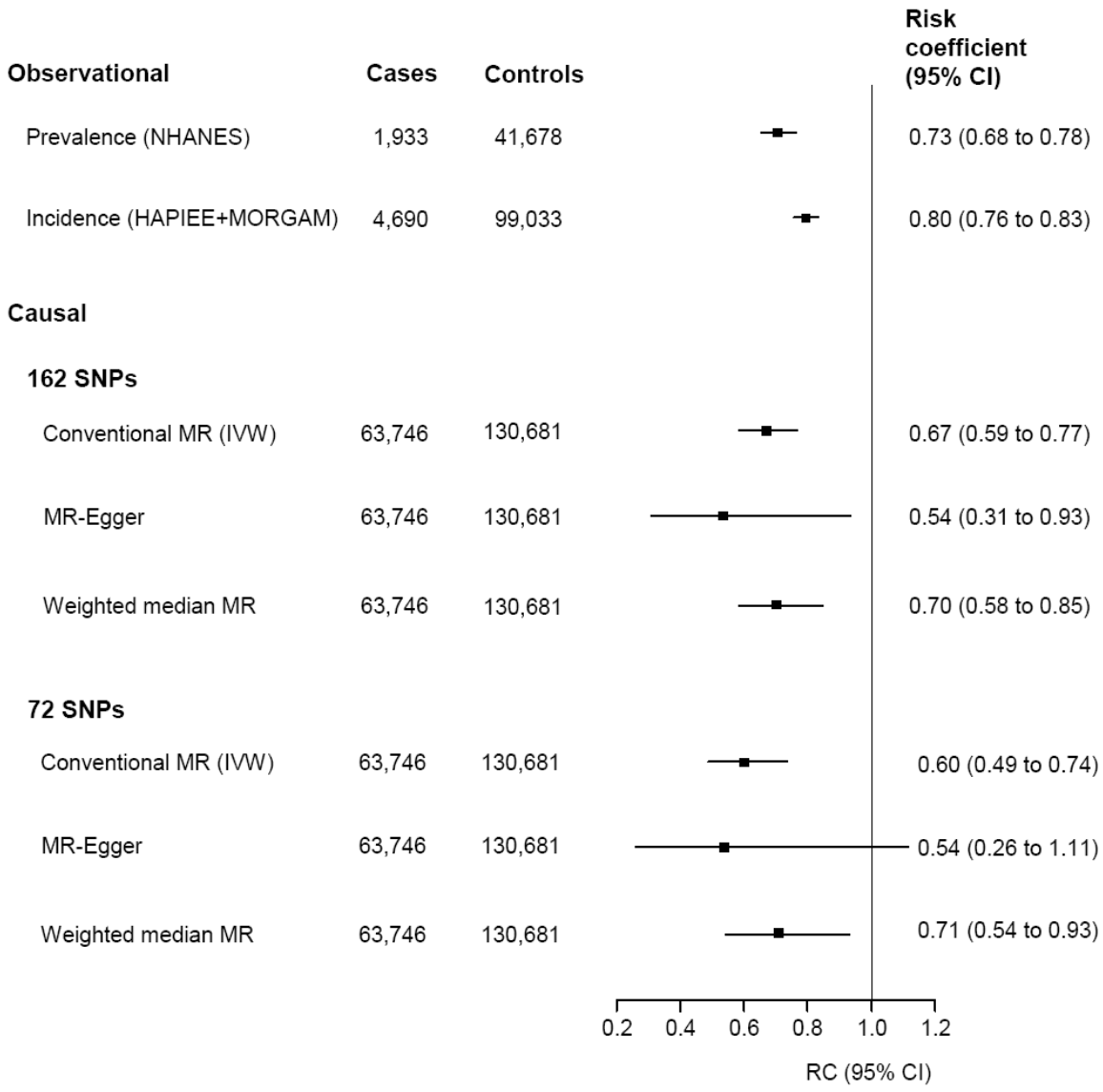


Figure 31. Comparison of observational and causal estimates for risk of coronary heart disease, per 3.6 years of educational attainment. The two observational estimates are provided according to prevalent and incident CHD cases, respectively. The risk coefficient (RC) for *observational incident* cases was derived by meta-analysis of hazard ratios from the HAPIEE and MORGAM studies. The RCs for the *observational prevalent* cases, and the six causal estimates from Mendelian randomisation (MR) are all odds ratios (see Supplementary Methods for a full description of each analysis). CHD, coronary heart disease. CI, Confidence Interval. NHANES, The National Health and Nutrition Examination Survey. HAPIEE, Health, Alcohol and Psychosocial factors In Eastern Europe. MORGAM, MOnica Risk, Genetics, Archiving and Monograph. IVW, Inverse-variance weighted approach.

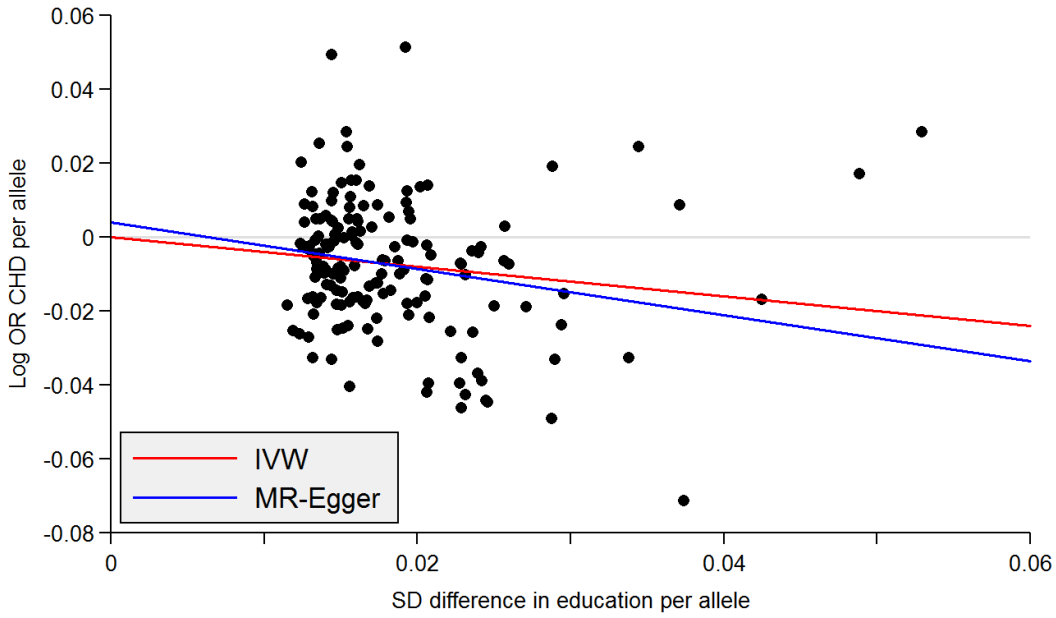


Figure 32. Scatter plot of 162 SNPs associated with education and their risk of CHD (with MR-Egger).

The red line shows causal regression estimates from conventional Mendelian randomization (MR), inverse variance weighted (IVW).
The blue line shows causal regression estimates from MR-Egger.
SNP, single nucleotide polymorphism. CHD, coronary heart disease. OR, odds ratio.

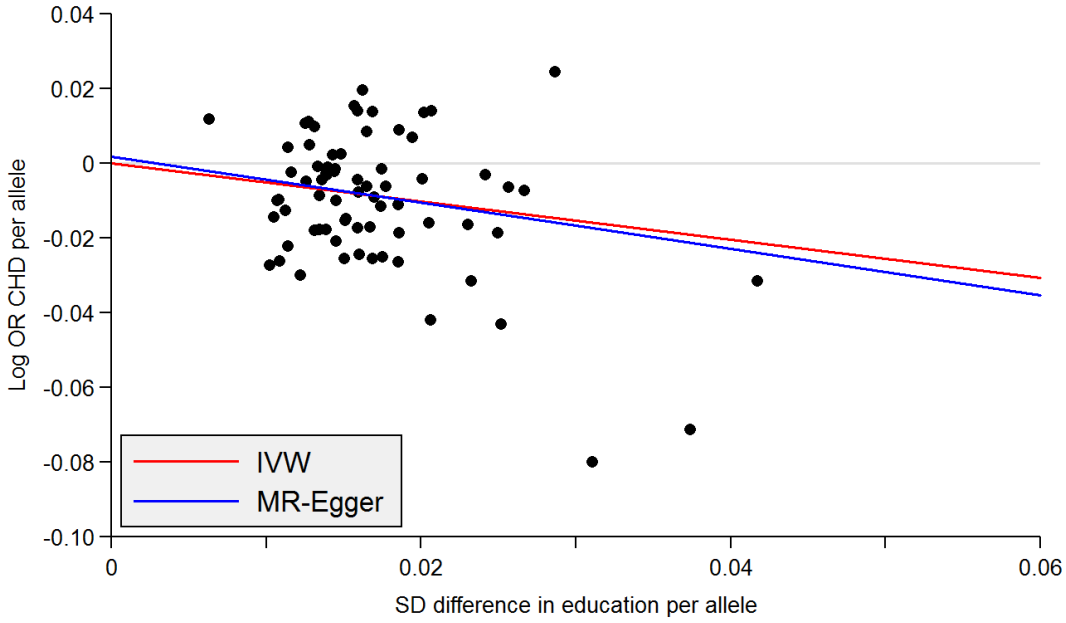


Figure 33. Scatter plot of the 72 SNPs associated with education and their risk of CHD (with MR-Egger).

The red line shows causal regression estimates from conventional Mendelian randomization (MR), inverse variance weighted (IVW).
The blue line shows causal regression estimates from MR-Egger.
SNP, single nucleotide polymorphism. CHD, coronary heart disease. OR, odds ratio.

Further sensitivity analyse are reported in Table 7. Briefly, an analysis, which can account for some measurement error in my genetic instruments for exposure (so-called MR-Egger + SIMEX)(238), gave similar findings to Standard MR-Egger, suggesting that any residual measurement error in the SNP-education estimate is unlikely to distort the main analysis.

Results from modal-based MR approaches were overall consistent with the hypothesis that unbalanced directional pleiotropy was not driving the conventional MR result. Specifically, the first *Mode-Based Estimate* yielded directionally concordant point estimates. However, this test was grossly underpowered to detect a causal effect. The second *Largest Heterogeneous Subset* analyses, by contrast, were better powered. The vast majority of SNPs (i.e. 90%) were highly homogeneous in their causal effect estimates. Removing these 0-10% heterogeneous SNPs made little difference to the point estimates. Furthermore, all these MR estimates were highly unlikely to have been observed by chance alone.

I also performed robustness checks by omitting SNPs with higher levels of missing data, as well as SNPs with proxy outcome data. These gave similar results in terms of direction, magnitude and statistical significance (Table 7). Collectively, all these sensitivity analyses make it less likely for presence of missing data or pleiotropic effects to have grossly biased my main causal analysis.

Analysis	Causal effect (OR) estimate for risk of CHD (95% CI)	Causal effect P-value
Set of 162 SNPs (I^2 statistic=0.661)		
Conventional MR (IVW)	0.67 (0.59 to 0.77)	2.9×10^{-8}
Weighted-Median MR	0.70 (0.58 to 0.85)	1.8×10^{-4}
Standard MR-Egger	0.54 (0.31 to 0.93)	0.029
Adjusted MR-Egger (+SIMEX)	0.41 (0.19 to 0.87)	0.022
Mode-Based Estimate	0.84 (0.44 to 1.60)	0.255
Largest Homogeneous Subset-MR (3 tests below):		
Minus 2 most heterogeneous SNPs = 160 SNPs (Heterogeneity P≥0.05)	0.68 (0.59 to 0.77)	1.8×10^{-8}
Minus 4 most heterogeneous SNPs = 158 SNPs (Heterogeneity P≥0.20)	0.66 (0.59 to 0.75)	8.8×10^{-10}
Minus 12 most heterogeneous *causal* SNPs = 150 SNPs (Heterogeneity P≥0.20)	0.75 (0.67 to 0.86)	1.8×10^{-5}
Minus (47 proxies) = 115 SNPs. Conventional MR estimate (IVW)	0.62 (0.52 to 0.73)	1.3×10^{-7}
Minus (21 with >10% missing data) = 141 SNPs. Conventional MR estimate (IVW):	0.70 (0.60 to 0.80)	1.4×10^{-6}
Set of 72 SNPs (I^2 statistic=0.934)		
Conventional MR (IVW)	0.60 (0.49 to 0.74)	6.2×10^{-6}
Weighted-Median MR	0.71 (0.54 to 0.93)	0.014
Standard MR-Egger	0.54 (0.26 to 1.11)	0.099
Adjusted MR-Egger (+SIMEX)	0.43 (0.17 to 1.12)	0.088
Mode-Based Estimate	0.78 (0.40 to 1.54)	0.490
Largest Homogeneous Subset-MR (3 tests below):		
Minus 2 most heterogeneous SNPs = 70 SNPs (Heterogeneity P≥0.05)	0.65 (0.53 to 0.79)	5.8×10^{-5}
Minus 4 most heterogeneous SNPs = 68 SNPs (Heterogeneity P≥0.20)	0.64 (0.53 to 0.78)	9.4×10^{-5}
Minus 5 most heterogeneous *causal* SNPs = 67 SNPs (Heterogeneity P≥0.20)	0.67 (0.55 to 0.81)	7.6×10^{-5}

Table 7. Sensitivity analyses of Mendelian Randomization estimates. *All causal effects are expressed as change in Odds Ratio of coronary heart disease (CHD), per 3.6 years (1-SD) of longer education. The adjusted MR-Egger regression estimates are the results of 10,000 simulations. SNP, single nucleotide polymorphism. MR, Mendelian randomization. IVW, inverse-variance weighted (analysis). SIMEX, simulation extrapolation.*

3.1.3. Reverse direction Mendelian randomization

I found little evidence for the hypothesis that genetic liability for CHD risk is associated with educational outcomes. Namely, 1-log greater genetic risk of CHD was associated with 0.2 (-1.3 to 1.6) days of longer educational attainment. Results were unchanged after applying MR-Egger and weighted-median MR (Figure 34). The results from individual SNPs are shown in Figure 35.

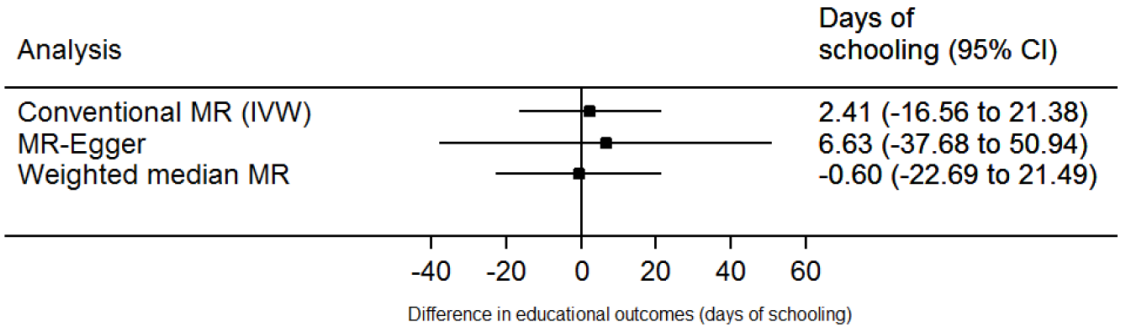


Figure 34. Association of genetic liability to CHD (exposure) on the numbers of days of schooling (outcome). *Causal estimates are expressed as difference in days of education, per 1-log unit increase in risk of coronary heart disease (CHD) as instrumented by 53 SNPs. CI, confidence interval. IVW, inverse variance weighted approach.*

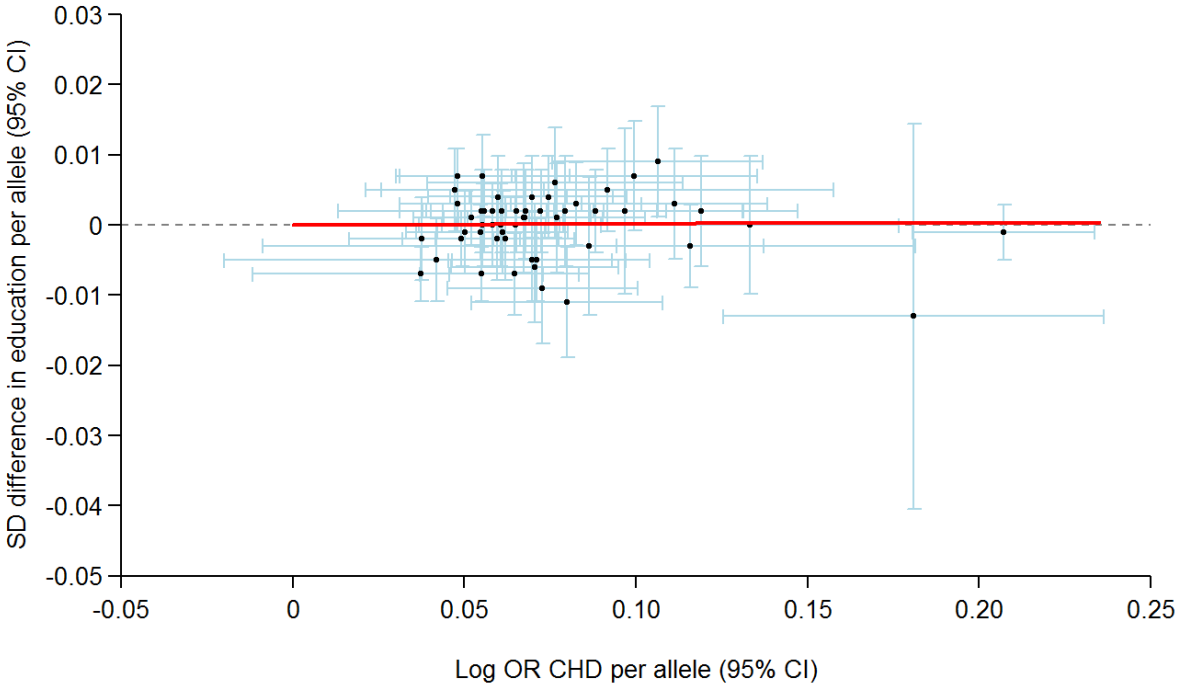


Figure 35. Scatter plot of the 53 SNPs associated with CHD development and their educational outcomes (with 95% confidence intervals). *The red line represents the regression slope of the causal effect estimate (derived by the inverse-variance weighted Mendelian randomization method). SNP, single nucleotide polymorphism. CHD, coronary heart disease. OR, odds ratio. CI, confidence interval.*

3.1.4. Mendelian randomization from education to CVD risk factors

To identify potential risk factors that could mediate the association between education and CHD, I investigated whether genetic predisposition towards higher education was associated with established cardiovascular risk factors. Table 8 shows that, in conventional MR analyses, a 1-SD longer education (due to genetic predisposition across 162 SNPs) was associated with a 35% lower odds of smoking, lower BMI (0.17 kg/m^2), lower triglycerides (0.14 mmol/l) and a greater HDL-cholesterol (0.15 mmol/l), with a P-value smaller than 0.001 for each of these four outcomes. Associations with diabetes and systolic blood pressure were in the anticipated direction, but these effects may have been due to chance and/or insufficient statistical power (P-values = 0.05 to 0.08).

3.1.5. Observational associations

Based on NHANES data, each additional 3.6 years of education (1-SD) was associated with 27% lower odds of prevalent CHD (odds ratio = 0.73 [95% confidence interval 0.68 to 0.78], illustrated in Figure 31 above). In prospective analyses, 3.6 years of additional education was associated with a 20% lower risk of incident CHD in the HAPIEE and MORGAM studies, with a pooled hazard ratio of 0.80 (0.76 to 0.83). Cohort-specific results from MORGAM are additionally shown in Figure 36.(245, 246) These observational estimates were robust to sensitivity analyses accounting for different case definitions, age at first CHD event, and confounding by other measures of socioeconomic position (Table 9). I also saw evidence for a dose-response relationship between the amount of education and risk of CHD (Figures 37 & 38).

Outcome	Causal effect (95% CI)	P-value
Binary traits		
Smoking status	**Odds Ratio = 0.65 (0.54 to 0.79)**	**0.001**
Diabetes (type 2)	Odds Ratio = 0.75 (0.56 to 1.01)	0.057
Continuous traits		
Systolic BP	-1.36 (-2.85 to 0.12) mmHg	0.075
Diastolic BP	-0.23 (-1.22 to 0.76) mmHg	0.645
LDL-cholesterol	-0.03 (-0.10 to 0.05) mmol/L	0.513
HDL-cholesterol	**0.15 (0.07 to 0.23) mmol/L**	**0.001**
Triglycerides	**-0.14 (-0.22 to -0.06) mmol/L**	**0.001**
Glucose	-0.02 (-0.08 to 0.03) mmol/L	0.441
BMI	**-0.17 (-0.26 to -0.08) kg/m^2**	**0.001**
Height	0.06 (-0.03 to 0.16) cm	0.208

Table 8. Causal effects from 3.6 years of education, to ten cardiovascular risk factors. *All analyses are based on a common set of 102 single nucleotide polymorphisms associated with education, available in 8 genome-wide association study consortia. Bold font denotes causal effect estimates with strong statistical evidence (p<0.001). Estimates are expressed per 1-standard deviation increase in years of education (equivalent to 3.6 years) as absolute values for continuous risk factors, and as odds ratios for binary traits. CI, confidence interval. LDL, low density lipoprotein; HDL, high density lipoprotein; BP, Blood Pressure; BMI, Body Mass Index.*

Case definition/ sub-analysis	Cases (*n*)	Controls (*n*)	Mean age at first event	Association with CHD, per 1-SD longer education	Used in figure 31?
Prevalence					
NHANES*					
Nonfatal CHD					
All ages	2,846	40,823	55.1	OR = 0.73 (0.68; 0.78)	Yes
All ages (no missing data, to compare with SES-adjusted estimate below)	1,234	16,790	55.1	OR = 0.75 (0.67; 0.83)	
All ages, fully SES-adjusted†	1,234	16,790	55.1	OR = 0.73 (0.62; 0.85)	
Age of first event <66y	1,907	40,823	48.7	OR = 0.72 (0.66; 0.78)	
Nonfatal AMI only					
All ages	1,933	41,678	54.7	OR = 0.71 (0.65; 0.77)	
Incidence					
HAPIEE‡					
Fatal / nonfatal CHD	632	22,879	65.0	HR = 0.75 (0.69; 0.81)	Yes
Nonfatal CHD only	338	23,138	64.1	HR = 0.81 (0.72; 0.91)	
Fatal CHD only	309	23,202	67.2	HR = 0.71 (0.64; 0.79)	
Fatal CVD	621	22,890	67.4	HR = 0.75 (0.70; 0.82)	
MORGAM‡					
Fatal / nonfatal CHD	6 522	90 526	63.2	HR = 0.83 (0.80; 0.86)	Yes

Table 9. Sensitivity analyses for observational estimates.SD, standard deviation. CHD, coronary heart disease. AMI, acute myocardial infarction.

** Adjusted for age and sex (additionally weighted to account for oversampled design, response rate, and geographical clustering).*

† Adjusted for age, sex, ethnicity, citizenship, country of birth, military service, marital status, household size, family income: poverty threshold ratio.

‡ Adjusted for age, sex, and country of survey.

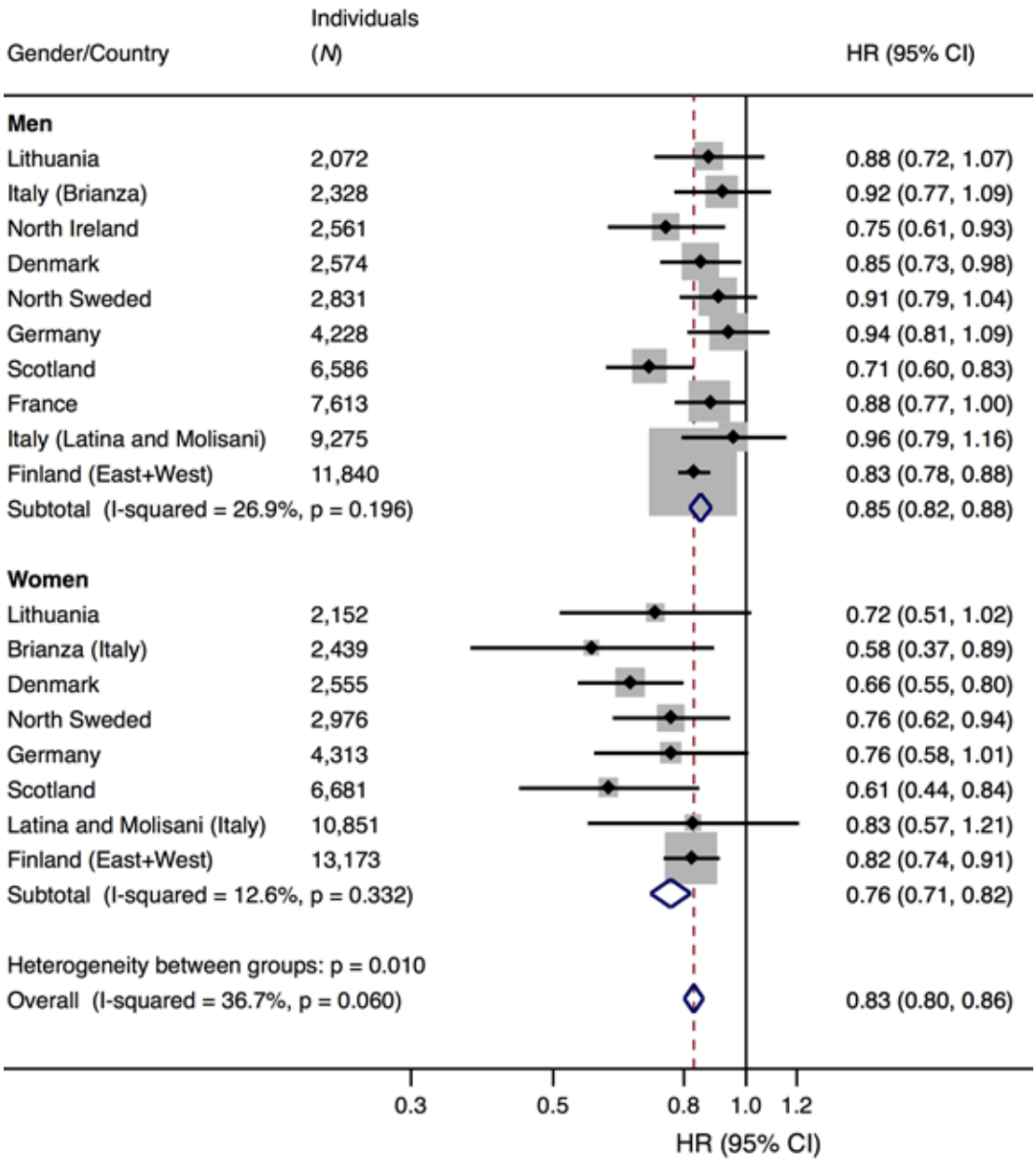


Figure 36. Observational estimates from the MORGAM consortium, showing cohort-level estimates and the results from meta-analysis. *Meta-analysis performed using inverse-variance weighted fixed-effect modelling. CHD, coronary heart disease. SD, standard deviation. HR, hazard ratio. CI, confidence interval.*

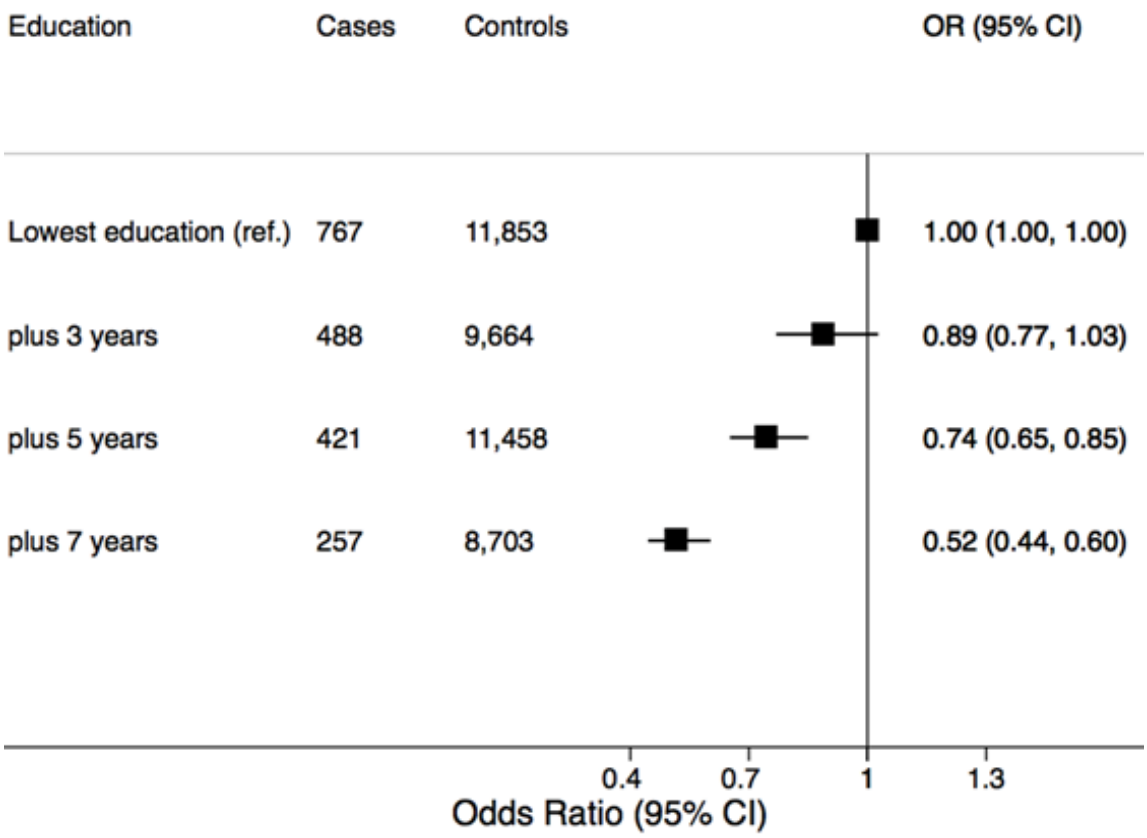


Figure 37. Dose response relationship between education and CHD prevalence, using observational NHANES data. *Lowest education group represents "some high school" in USA system, i.e. typically 16-year old pupil. Logistic regression model was adjusted for age and sex. CHD, coronary heart disease; NHANES, National Health and Nutrition Examination Survey; OR, odds ratio CI, confidence interval.*
P for trend < 0.0001.

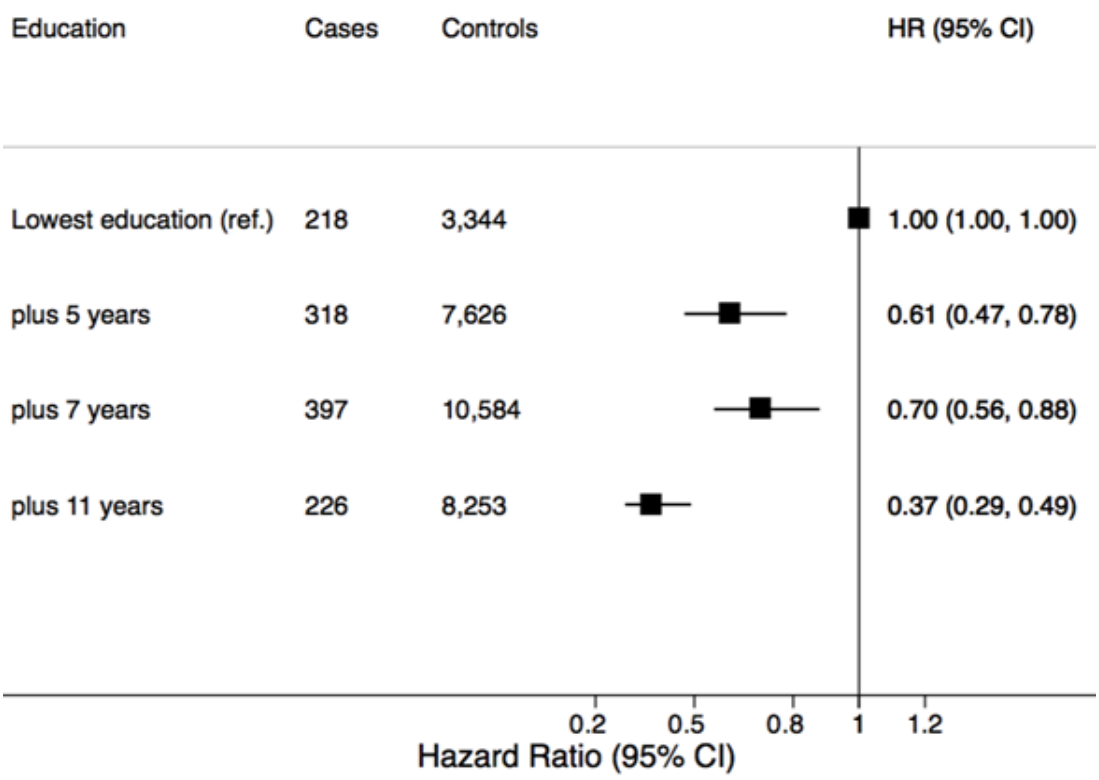


Figure 38. Dose response relationship between education and CHD incidence, using observational HAPIEE data. *Lowest education group represents "Primary education or lower", i.e. max. 4 years of education. Cox proportional hazard regression model was adjusted for age, sex and country. CHD, coronary heart disease; HAPIEE, The Health, Alcohol and Psychosocial factors In Eastern Europe Study; HR, hazard ratio CI, confidence interval. P for trend<0.001.*

3.2. Mediation and international differences

Section 3.2 will discuss results from the mediation analysis, as well as investigation of the causes of international differences. As a reminder, literature to date has been limited around to what extent psychosocial factors could mediate the presumably causal effect from socioeconomic factors to CVD. Secondly, literature to date has not explored whether psychosocial and socioeconomic factors could explain why CVD mortality is higher in ex-USSR countries, when compared to countries from Central Europe.

3.2.1. Baseline data

Baseline characteristics of participants in the analytical sample are shown in table 10. Participants who subsequently died of CVD had higher levels of most risk factors, when compared to those who did not die of CVD. There was an interaction between sex and country: men in Russia had higher hazard ratios than expected based on the sum of the “male” and the “Russian” indicator variables alone (HR for the interaction term=1.77 [1.23-2.55], p=0.002). This term was kept in all subsequent models.

	Total sample		% missing & imputed
	n / mean	% / SD	
Participants	20 867	100%	
Follow-up years (median, max)	7.2,	11.3	
Age (per year)	57.2	7.03	0%
Male	9 700	46%	0%
Conventional CVD risk factors			
Diabetes	1 554	7.4%	0.1%
Smoking Status:			
Never-smoker	9 941	48%	
Occasional/Past smoker	5 065	24%	
Daily smoker, 1-10 cigarettes/day	2 074	9.9%	0.6%
Daily smoker, 11-20 cigarettes/day	3 039	15%	
Daily smoker, >20 cigarettes/day	747	3.6%	
Blood pressure, systolic (mmHg)	138.9	22.5	9.2%
Total cholesterol (mmol/L)	5.95	1.25	10.5%
HDL cholesterol (mmol/L)	1.48	0.46	10.6%
Body Mass Index (kg/m^2)	28.0	4.85	9.2%
Physically inactive	1 666	8.0%	1.1%
Alcohol intake:			
Nil	4 033	19%	
Up to UK guidelines	12 389	59%	
Exceeding UK guidelines (1-2x over)	2 314	11%	1.2%
Exceeding UK guidelines (>2x over)	2 131	10%	
Alcohol drinking frequency:			
Non-drinker	3 942	19%	1.4%
< once/week	10 534	50%	
≥ once/week	6 391	31%	
Binge drinking (≥ once/month)	2 705	13%	1.4%
Possible problem drinking (CAGE ≥ 2)	1 408	6.7%	8.1%

Psychological factors			
Marital Status:			
Married/cohabiting	15 713	75%	0.2%
Divorced/widowed	4 257	20%	
Single	897	4.3%	
Social Support			
Contact with relatives <once/month	4 966	24%	0.4%
Contact with friends <once/month	7 554	36%	0.6%
Not a member of a club	17 487	84%	0.6%
Depression symptoms (possible case)	4 612	22%	11.9%
Low perceived control (SD scale)	0.00	1.01	1.5%
Socioeconomic factors			
Education			
Tertiary	5 263	25%	0.2%
Secondary	13 459	65%	
Primary	2 143	10%	
Material possessions			
Low amenities, current (SD scale)	0.00	1.00	2.6%
Low amenities, early life (SD scale)	0.00	1.01	2.9%
Deprivation, current (SD scale)	0.00	1.02	0.9%
Deprivation, early life (SD scale)	0.00	0.99	1.0%
Unemployment, current	897	4.3%	0.4%
Unemployment, long term	1 690	8.1%	1.8%
Change in status since 1989:			
Improved	5 071	24%	0.9%
Stayed the same	10 079	48%	
Declined	5 718	27%	

Table 10a. Baseline characteristics of the analytical sample.
The sample shown is multiply imputed, combined across three countries and two genders. The final column shows amount of missing/imputed data among the total sample. SD=Standard Deviation.

	Czech Republic		Poland		Russia	
	n / mean	% / SD	n / mean	% / SD	n / mean	% / SD
Participants	6 905	33%	7 039	34%	6 923	33%
Follow-up years (median, max)	9.6,	11.3	7.1,	8.9	6.6,	8.0
Events (CVD mortality)	173	31%	134	24%	249	45%
Conventional risk factors						
Age, mean (SD)	57.6	7.1	56.8	7.0	57.3	7.0
Male	3 130	45%	3 460	50%	3 110	45%
Diabetes	677	10%	629	9%	248	4%
Smoking Status:						
Non-smoker	3 119	45%	2 805	40%	4 019	58%
Occasional/Past smoker	2 088	30%	2 051	29%	926	13%
Daily smoker, 1-10/day	746	11%	649	9%	679	10%
Daily smoker, 11-20/day	824	12%	1 167	17%	1 049	15%
Daily smoker, >20 /day	129	2%	368	5%	250	3.6%
Blood pressure, systolic (mmHg)	138.4	20.3	137.0	21.8	141.3	24.3
Cholesterol,total (mmol/L)	5.7	1.1	5.8	1.3	6.3	1.3
HDL (mmol/L)	1.4	0.4	1.5	0.4	1.6	0.5
Body Mass Index (kg/m^2)	28.0	4.6	27.8	4.6	28.2	5.3
Physically inactive	827	12%	572	8%	269	4%
Alcohol intake:						
Nil	789	11%	2 273	32%	971	14%
Up to UK guidelines	3 937	57%	3 335	47%	5 117	74%
Exceeding UK guidelines (1-2x over)	1 137	16%	670	10%	507	7%
Exceeding UK guidelines (>2x over)	1 042	15%	760	11%	328	4%
Alcohol drinking frequency:						
Non-drinker	802	12%	2 206	31%	934	13%
< once/week	3 243	47%	3 106	44%	4 184	60%
≥ once/week	2 859	41%	1 727	25%	1 805	26%
Binge drinking (≥1/month)	968	14%	458	7%	1 279	19%
Possible problem drinking (CAGE ≥2)	366	5%	348	5%	694	10%

Psychosocial factors						
Marital Status:						
Married/cohabiting	5 244	76%	5 429	77%	5 042	73%
Divorced/widowed	1 469	21%	1 206	17%	1 586	23%
Single	191	3%	403	6%	295	4%
Social Support						
Contact with relatives <once/month	757	11%	2 192	31%	2 008	29%
Contact with friends <once/month	1 730	25%	2646	38%	3 186	46%
Not a member of a club	4 992	72%	6 167	88%	6322	91%
Depression symptoms (possible case)	1 253	18%	1 642	23%	1 715	25%
Low perceived control (SD scale)	0.17	1.00	-0.17	0.99	0.02	0.98
Socioeconomic factors						
Education						
Tertiary	1 016	15%	2 206	31%	2 042	29%
Secondary	5 089	74%	4 157	59%	4 214	61%
Primary	800	12%	677	10%	667	9.6%
Material possessions						
Low amenities, current (SD scale)	-0.23	1.00	-0.07	0.98	0.30	0.95
Low amenities, early life (SD scale)	-0.48	0.75	-0.09	1.03	0.54	0.93
Deprivation, current (SD scale)	-0.31	0.75	0.10	0.97	0.41	1.15
Deprivation, early life (SD scale)	0.04	0.93	-0.14	0.92	0.10	1.09
Unemployment, current	194	3%	389	6%	313	4.5%
Unemployment, long term	398	6%	562	8%	735	11%
Change in status since 1989:						
Improved	2 244	33%	1 765	25%	1051	15%
Stayed the same	3 695	54%	3 274	47%	3113	45%
Declined	965	14%	1 997	28%	2758	40%

Table 10b. Baseline characteristics of the analytical sample, stratified by country.

The sample shown is multiply imputed, combined across two genders.
HDL=High Density Lipoproteins.
SD=Standard Deviation.

3.2.2. Independent associations with CVD mortality

As expected, conventional CVD risk factors were associated with CVD mortality (Figure 38).

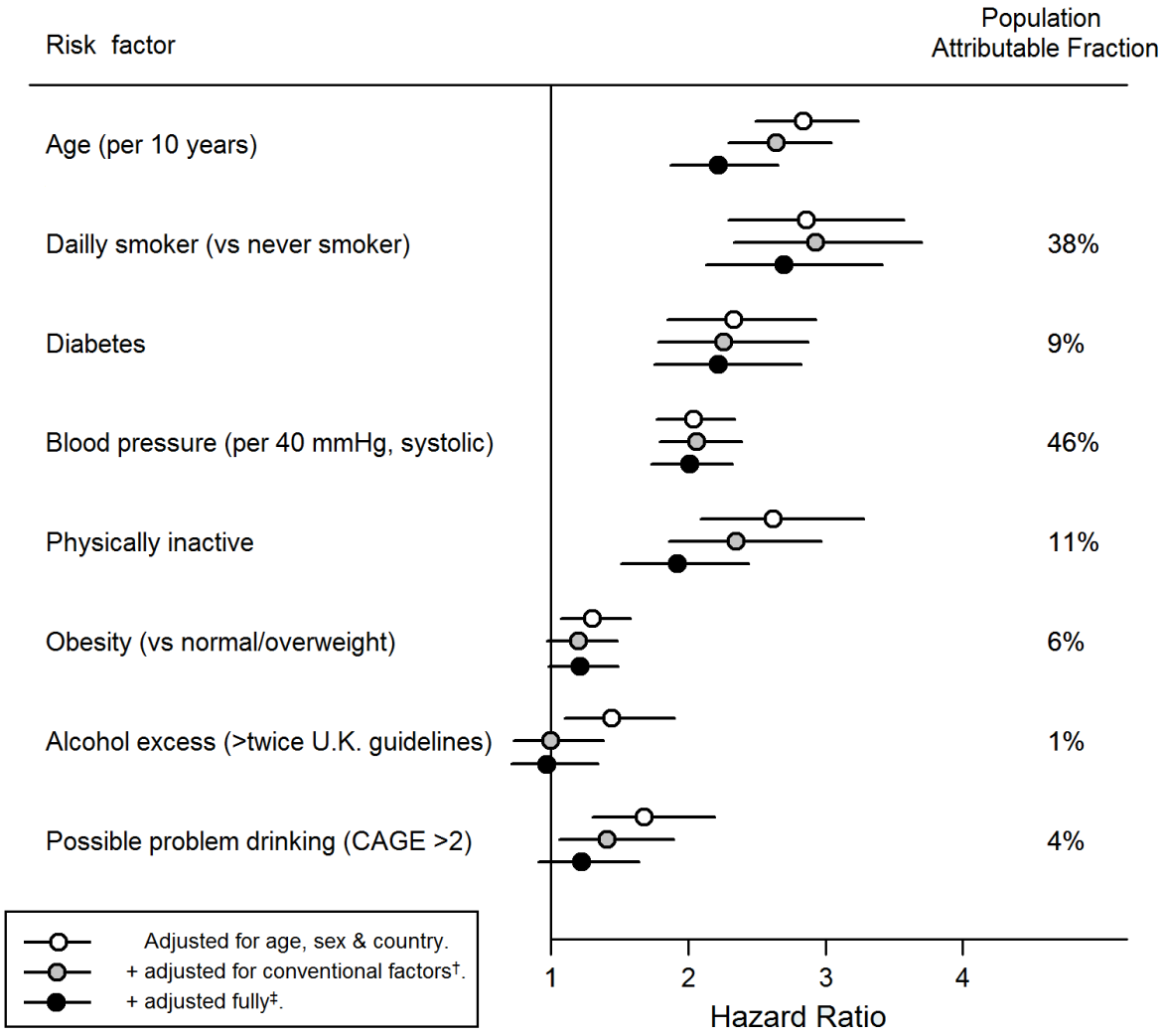


Figure 38. Associations of conventional cardiovascular risk factors with cardiovascular mortality.

†Conventional factors = diabetes, smoking, blood pressure, cholesterol, HDL, BMI, physical activity.

*‡Full adjustment = conventional factors + material possessions, depression, seeing relatives, seeing friends, friends*gender interaction, marital status, unemployment.*

Associations between 14 psychosocial and socioeconomic exposures and CVD mortality are subsequently tested. In model 1, 13 out of 14 factors tested were associated with CVD mortality, with HRs ranging from 2.96 (1.97-4.46, $p<0.0001$) for *current unemployment* to HR= 1.14 (1.05-1.23, $p=0.012$) per one standard deviation increase in *early life deprivation*. Twelve associations remained significant after adjustment for eleven conventional CVD risk factors in model 2; this attenuated the remaining hazard ratios by around a quarter.

Infrequent contact with friends was associated with outcomes more strongly in women than in men (Model 1 interaction with sex HR=1.83 [1.26-2.66], $p=0.002$, thus satisfying Bonferroni criteria). Model 3 therefore included both the conventional binary variable *"low friends"* (which was handled similarly to the other five psychosocial and socioeconomic factors), as well as the interaction term "*gender*low friends*" (interactions were not used for the remaining five factors). There was no evidence of effect modification by country (data not shown).

Following full adjustment for other psychosocial and socioeconomic factors in model 3, 6 factors remained associated with the outcome ($0.00001 < p < 0.007$): *depression, low material amenities*, *current unemployment, infrequent contact with relatives*, *infrequent contact with friends (for female participants only), and single marital status* (Figure 39). Following this full adjustment, the hazard ratio for *education* (HR = 1.11 [0.96-1.29] per 3 years of additional education) was largely attenuated in comparison with age-sex adjusted models. Consequently, I did not consider education to be one of the core variables in subsequent multivariate analyses. The population attributable fractions (PAF), in models adjusted for conventional risk factors ranged from around 9% (for *current unemployment, infrequent contact with relatives*) to 22% (for *low material amenities*). Test of effect modification gave similar results in analyses stratified by gender or cohort (data not shown). Sensitivity analyses gave similar results when limiting follow-up time to 8 years in all

three countries; excluding those participants with less than 2 years of follow-up; or when excluding imputed data (data not shown).

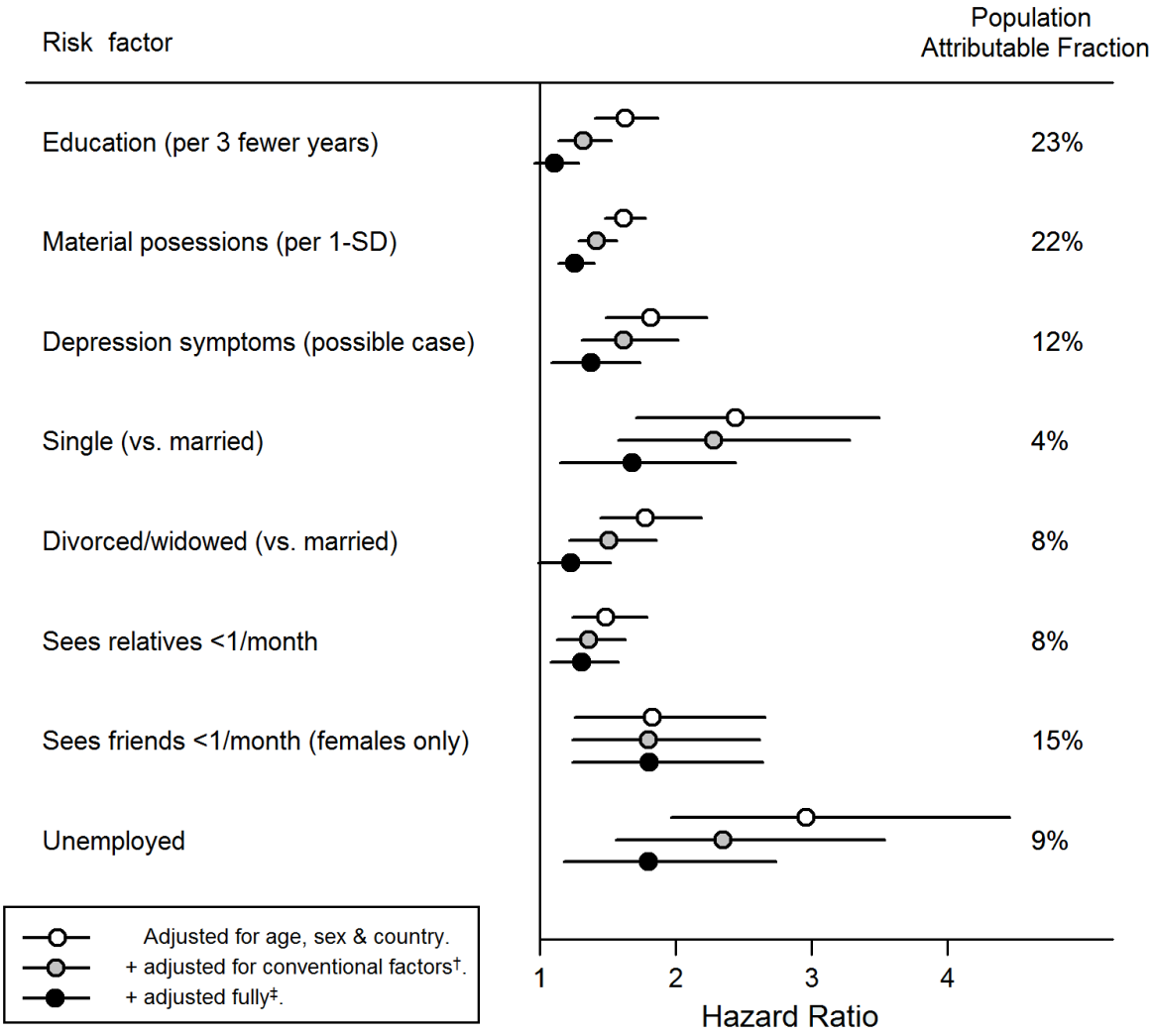


Figure 39. Associations of psychosocial and socioeconomic factors with cardiovascular mortality.

†Conventional factors = diabetes, smoking, blood pressure, cholesterol, HDL, BMI, physical activity.

*‡Full adjustment = conventional factors + material possessions, depression, seeing relatives, seeing friends, friends*gender interaction, marital status, unemployment.*

3.2.3. Attenuation and mediation

I examined to what degree the hazard ratios associated with the psychosocial factors attenuated following various adjustments (Figure 40). Adjustment for eleven conventional risk factors attenuated these by about one quarter. Additional adjustment for other psychosocial factors attenuated these by an additional quarter. As two exceptions, the hazard ratios for limited contact with friends and relatives did not attenuate to this level. This suggests very little overlap between the potential effects of these facets of social support, and that conventional risk factors may play a negligible role in mediating any of these effects. Overall, this data suggests that depression and social support constructs, if causal, may operate independently to socioeconomic constructs, potentially along separate mechanistic pathways.

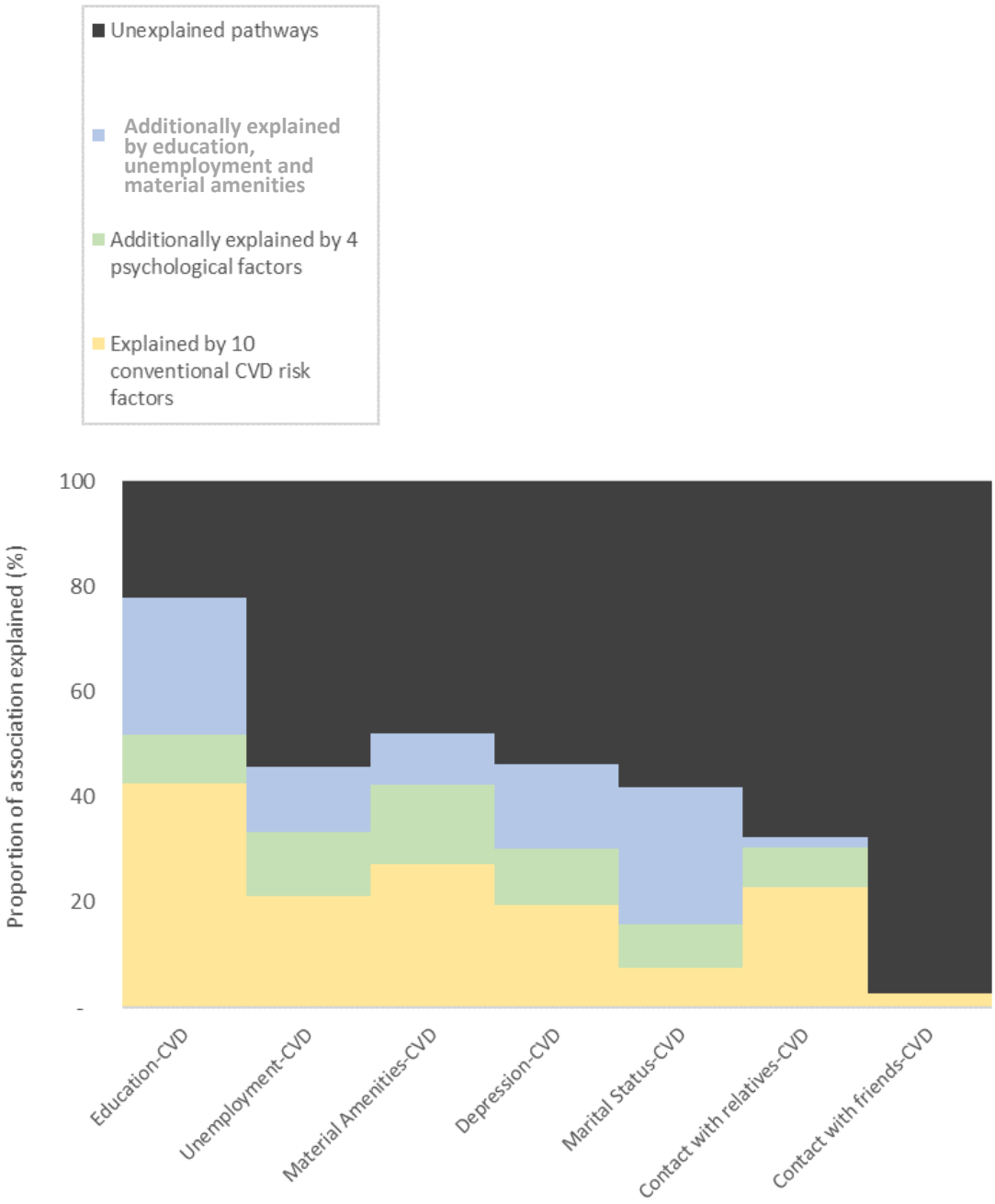


Figure 40. Attenuation among three socioeconomic (left side) and four psychosocial (right side) predictors of cardiovascular mortality. *Created from four sequentially-adjusted models. The total height of each predictor (i.e. 100%) is equivalent to its association with cardiovascular mortality in the simple age- and sex-adjusted model 1. The yellow area represents the subsequent attenuation in association, following additional adjustment for eleven conventional CVD risk factors in model 2 (equivalent to the yellow arrow in my causal diagram in figure 14). The green area represents the subsequent attenuation, following additional adjustment for four psychosocial variables (equivalent to the green arrow in figure 14). The blue area represents subsequent attenuation, following additional adjustment for two socioeconomic variables in model 3. The black area accounts for the unexplained part still present in model 3 (i.e. direct or non-mediated effects, equivalent to the black arrow in figure 14).*

As few studies have reported to what degree the associations between alcohol and mortality attenuate after adjustment for psychosocial factors, I additionally report these here. There was little evidence that total alcohol consumption or binge drinking was associated with CVD mortality in models adjusted for conventional risk factors. However, people scoring positive on the CAGE screening questionnaire for possible problems with alcohol had 41% greater risk of CVD in crude models. This attenuated by around one half following adjustment for six psychosocial factors. This observation can be interpreted in two ways. First, some of the previously reported associations between alcohol and mortality in the literature may have been biased away from the null, since they did not control comprehensively enough for psychosocial factors which act as confounders. Second, it may instead be the case that alcohol excess causes psychosocial stress which causes mortality, and therefore psychosocial factors may be causal mediators. The existing literature does not appear to be developed enough to firmly inform which of these two explanations are more likely to be true.

3.2.4. International differences

CVD mortality risk was substantially higher in the Russian cohort. In analyses done among men only, the age-adjusted hazard ratio for being in Russia vs. Central Europe was 2.86 [2.31-3.54]; this excess risk was not reduced following adjustment for conventional CVD risk factors (HR=2.78 [2.15-3.59]), or following adjustment for conventional and psychosocial factors (HR=2.77 [2.13-3.61]). In analyses done among women only, the hazard ratio of being in Russia vs. Central Europe was 1.59 [1.15-2.19]; this difference was exacerbated following adjustment for conventional factors (HR=2.15 [1.45-3.18]) but returned to a level comparable to crude models following additional adjustment for psychosocial factors (HR=1.64 [1.09-2.46]). Similar results emerged in analyses where data from men and women were pooled together, and an explicit interaction was modelled between gender and country (Figure 41).

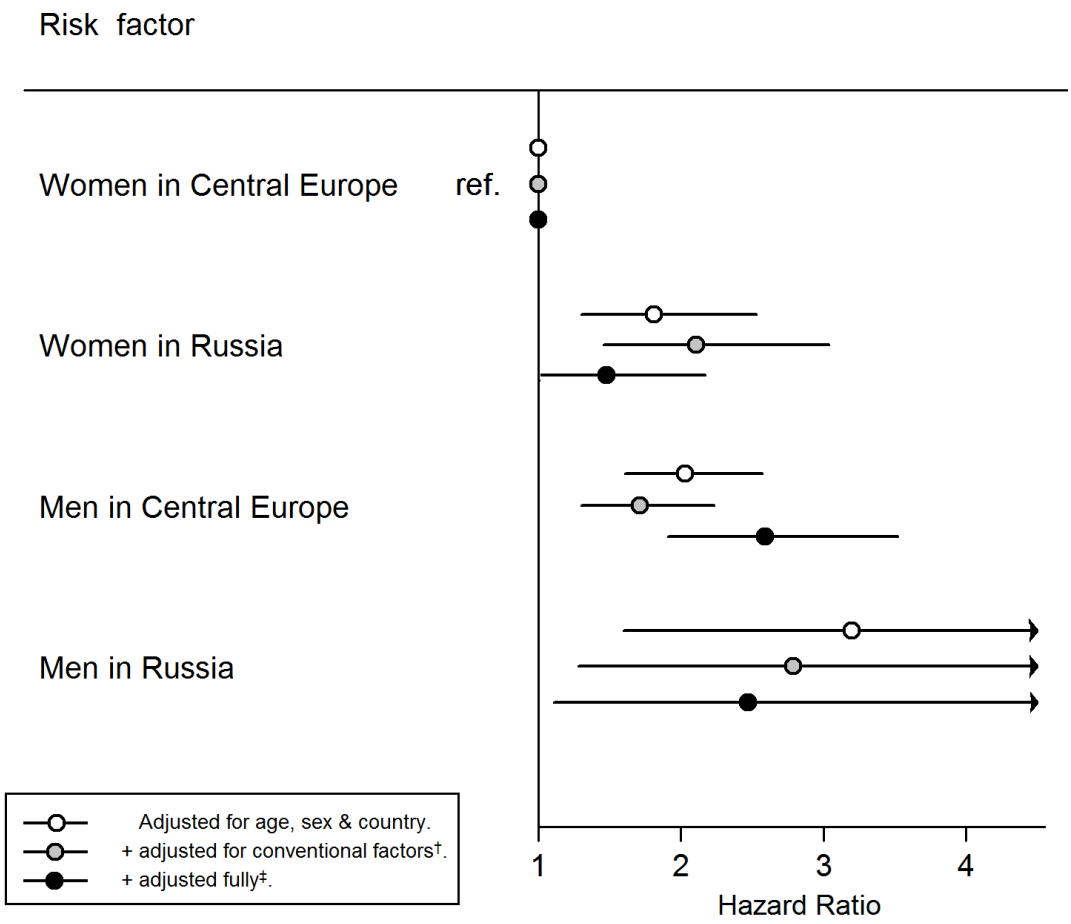


Figure 41. Associations of male sex and being in Russia (vs. being female in Central Europe) with cardiovascular mortality. *There were 109, 67, 198 and 175 events respectively in the four subgroups shown (from top to bottom).*

†Conventional factors = diabetes, smoking, blood pressure, cholesterol, HDL, BMI, physical activity, alcohol.

*‡Full adjustment = conventional factors + material possessions, depression, seeing relatives, seeing friends, friends*gender interaction, marital status, unemployment.*

Central Europe = Poland or Czech Republic.

3.3. Prediction

Section 3.3 will describe the results from my work on risk prediction. As a reminder, current cardiovascular risk prediction models available in Eastern Europe have substantial limitations. The first aim was to fit a new risk prediction model, using conventional CVD risk factors, and contemporary cohort data using the HAPIEE study (known as *Model 2* or *EastEur-SCORE*). A related aim was to externally evaluate how well this worked in the Estonian BioBank study.

The second weakness in the current evidence is how it is unclear whether the addition of psychosocial and socioeconomic risk factors can improve the performance of clinical risk prediction models worldwide. To investigate this, I added new predictors to my newly created Model 2 (EastEur-SCORE), to derive model 3 (known as *HAPIEE-SCORE*). I then evaluated the performance of HAPIEE-SCORE in external data from the Estonian BioBank study.

3.3.1. Data and model description

The derivation and validation samples are described in table 11 and figure 42. The validation dataset was about 3 times smaller than the derivation dataset, which methodologically may be considered an appropriate balance. While the male-female split was almost equal in the derivation data, there were only 34% males in the validation data. In terms of age profiles, the validation cohort spanned the whole distribution of the derivation cohort, but additionally included younger participants (aged 37 to 44) for whom predicted risk will be extrapolated. The validation cohort appeared to be slightly healthier in some respects (e.g. smoking, blood pressure, employment), but slightly less healthy in other respects (e.g. depression, single marital status, physical inactivity). Table 12 shows details of the two newly derived models (baseline risk, and beta coefficients from Hazard Ratios for each risk factor). Both new models were fully fitted, i.e. fully adjusted for all the risk factors, for which coefficients are shown.

Dataset	Derivation		Validation	
Country	Poland, Czech Republic, or Russia		Estonia	
Participants, N	14,598		4,632	
Follow-up, median years	7.2		8.3	
CVD mortality events (%)	338	(2.3%)	91	(2.0%)
SCORE risk factors				
Male, N (%)	6,910	(47%)	1,563	(34%)
Age, mean	57 (range 44 to 74)		51 (range 37 to 74)	
Diabetes (%)	953	(6.5%)	294	(6.3%)
Smoking Status				
never (%)	6,853	(47%)	2,652	(57%)
ex/light (%)	3,615	(25%)	699	(15%)
current (%)	4,130	(28%)	1,281	(28%)
Blood pressure (mmHg)	139		129	
Total cholesterol (mmol/L)	6.0		6.0	
Novel risk factors				
Body Mass Index	27.9		27.2	
Physically inactive (%)	1,057	(7.2%)	2,480	(54%)
Antihypertensive use	3,805	(26%)	1,163	(25%)
Education				
tertiary (%)	3,874	(27%)	1,147	(25%)
secondary (%)	9,509	(65%)	3,423	(74%)
primary or less (%)	1,215	(8%)	62	(1.3%)
Employment Status:				
Employed (%)	8,425	(58%)	3,714	(80%)
Unemployed (%)	679	(4.7%)	102	(2.2%)
Retired (%)	5,494	(38%)	816	(18%)
Marital Status:				
Married/cohabiting (%)	11,129	(76%)	2,544	(55%)
Divorced/widowed (%)	2,843	(20%)	1,586	(34%)
Single (%)	626	(4.3%)	502	(10.8%)
Depression, suspected (%)	2,979	(20%)	1,530	(33%)

Table 11. Baseline characteristics of the analytical sample.
The sample shown are complete cases, combined across two genders.

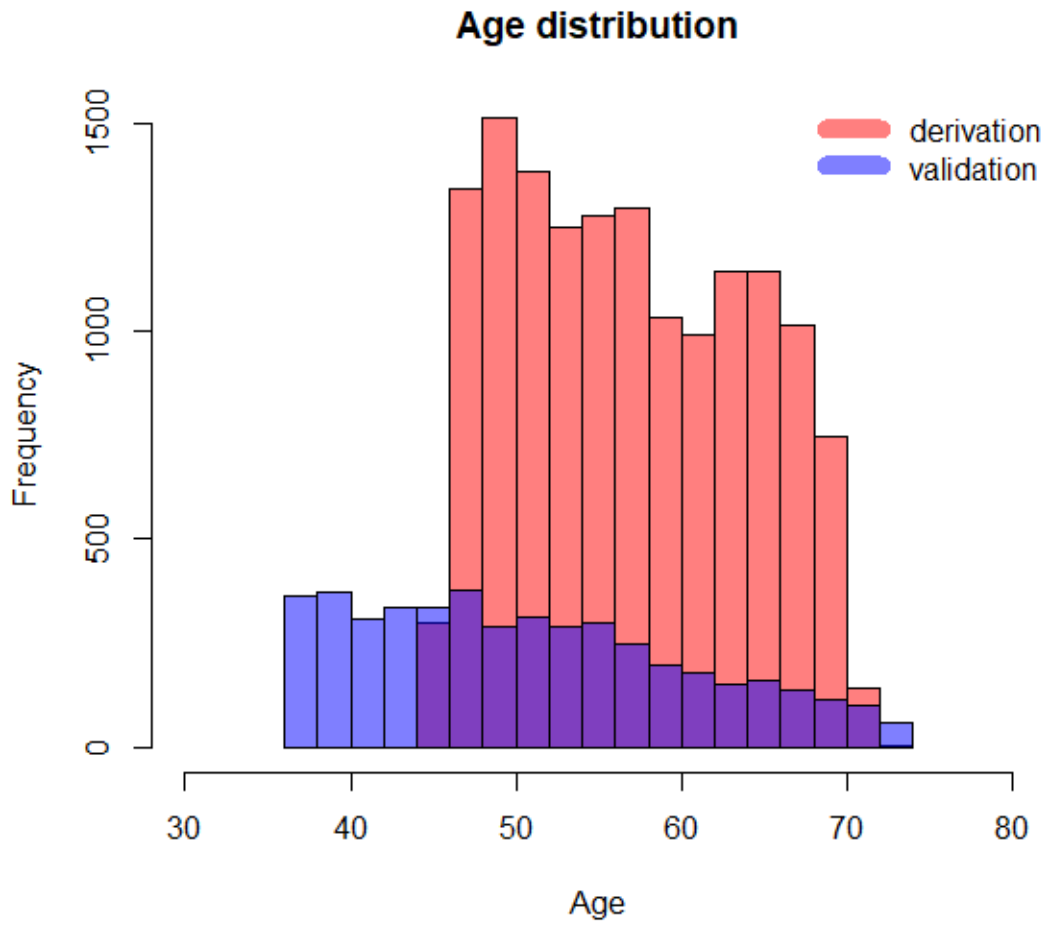


Figure 42. Age distributions of the derivation and validation samples.

	Name of new model	
	EastEur-SCORE	**HAPIEE-SCORE**
Baseline risk	0.002455	0.0009315
SCORE risk factors		
Male	1.457 (1.327 to 3.114)	1.940 (1.381 to 2.725)
Age (per 1 year)	1.100 (1.080 to 1.119)	1.092 (1.068 to 1.116)
Cholesterol, total (per 1 mmol/L)	1.028 (0.932 to 1.135)	1.063 (0.962 to 1.176)
Cholesterol^2 (centred at 6 mmol/L)	1.041 (1.019 to 1.062)	1.037 (1.016 to 1.059)
Smoking status		
Non smoker	ref.	ref.
Ex/light smoker	1.590 (1.145 to 2.207)	1.544 (1.112 to 2.143)
Current smoker	3.483 (2.604 to 4.660)	3.081 (2.283 to 4.158)
Blood pressure, systolic (per 1 mmHg)	1.017 (1.013 to 1.022)	1.020 (1.014 to 1.025)
Diabetes	2.302 (1.678 to 3.158)	2.226 (1.610 to 3.076)
High-risk area	2.033 (1.327 to 3.114)	1.803 (1.163 to 2.795)
High-risk are*Male interaction	1.581 (0.968 to 2.584)	1.964 (1.189 to 3.244)
Novel risk factors		
BMI (per 1-unit kg/m^2)	-	0.956 (0.919 to 0.996)
BMI^2 (centred at 23 kg/m^2)	-	1.005 (1.003 to 1.007)
Physical Inactivity (<150 min./week)	-	1.572 (1.135 to 2.178)
Antihypertensive use		1.243 (0.772 to 2.004)
Antihypertensives *blood pressure interaction		0.992 (0.983 to 1.002)
Educational attainment†	-	1.213 (0.998 to 1.474)
Employment status	-	
Employed	-	ref.
Unemployed	-	2.17 (1.321 to 3.564)
Retired	-	1.386 (1.044 to 1.84)
Marital Status	-	
Married/cohabiting	-	ref.
Divorced/widowed	-	1.482 (1.131 to 1.942)
Single	-	2.662 (1.701 to 4.165)
Depression (possible)	-	1.676 (1.299 to 2.161)

Table 12. Parameters of the two newly created models.

† = Education was treated as a categorical variable with three categories, and assuming a linear relationship across the three categories. The reference category was "tertiary education or above", which was compared against "secondary education" and "primary or less education" categories.

3.3.2. Calibration

Calibration plots are shown in figures 43 and 44. In most of the six panels, most of the blue dots follow and approximately straight line of best fit (line not shown). This lack of curvature suggests that model calibration at the low end of the risk spectrum is comparable to model calibration at the high end of the risk spectrum. Each of the plots are now discussed in turn, beginning with figure 43 (for derivation data), going from top to bottom.

Good calibration would be seen if some blue dots that appear above the grey diagonal line are compensated by some blue dots below the grey diagonal line. However, there appear to be some deviations from this in some of the panels, suggesting minor imperfections with calibration. The top panel suggests that the original SCORE model consistently overpredicted risk (across the entire risk distribution), in the derivation HAPIEE dataset. As expected, the newly derived models 2 and 3 appear better calibrated to the derivation data. This is typical and may reflect either a good model, or simply overfitting the derivation data. Hence calibration plots are most informative when looking at the externally validated data. These are shown in figure 44, and I will discuss these going from top to bottom.

In the validation data, the original SCORE model was better calibrated (with some over prediction seen among the highest-risk groups, seen in the top right corner). When looking at the middle right panel, it appears that newly derived EastEur-SCORE model 2 partly overpredicted risk in the validation dataset, and after adding psychosocial factors to model 3 then this over prediction became much larger (bottom panel). The pattern seen in these two panels is often seen in external validation studies, where newly derived models tend to over predict risk among those of highest risk. Such problems can sometimes be compensated by performing coefficient shrinkage, which was not done in my analyses.

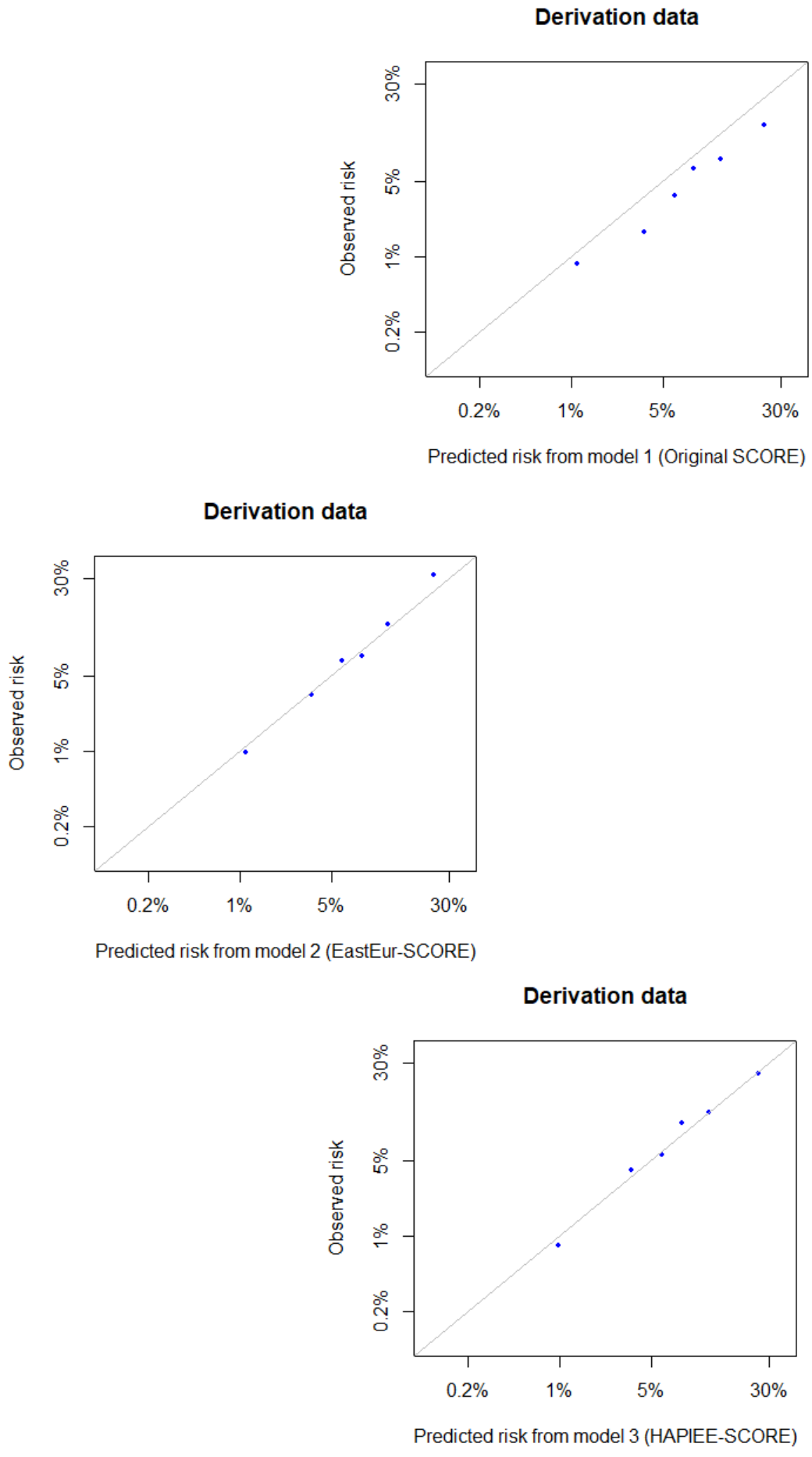


Figure 43. Calibration plots in the derivation data.

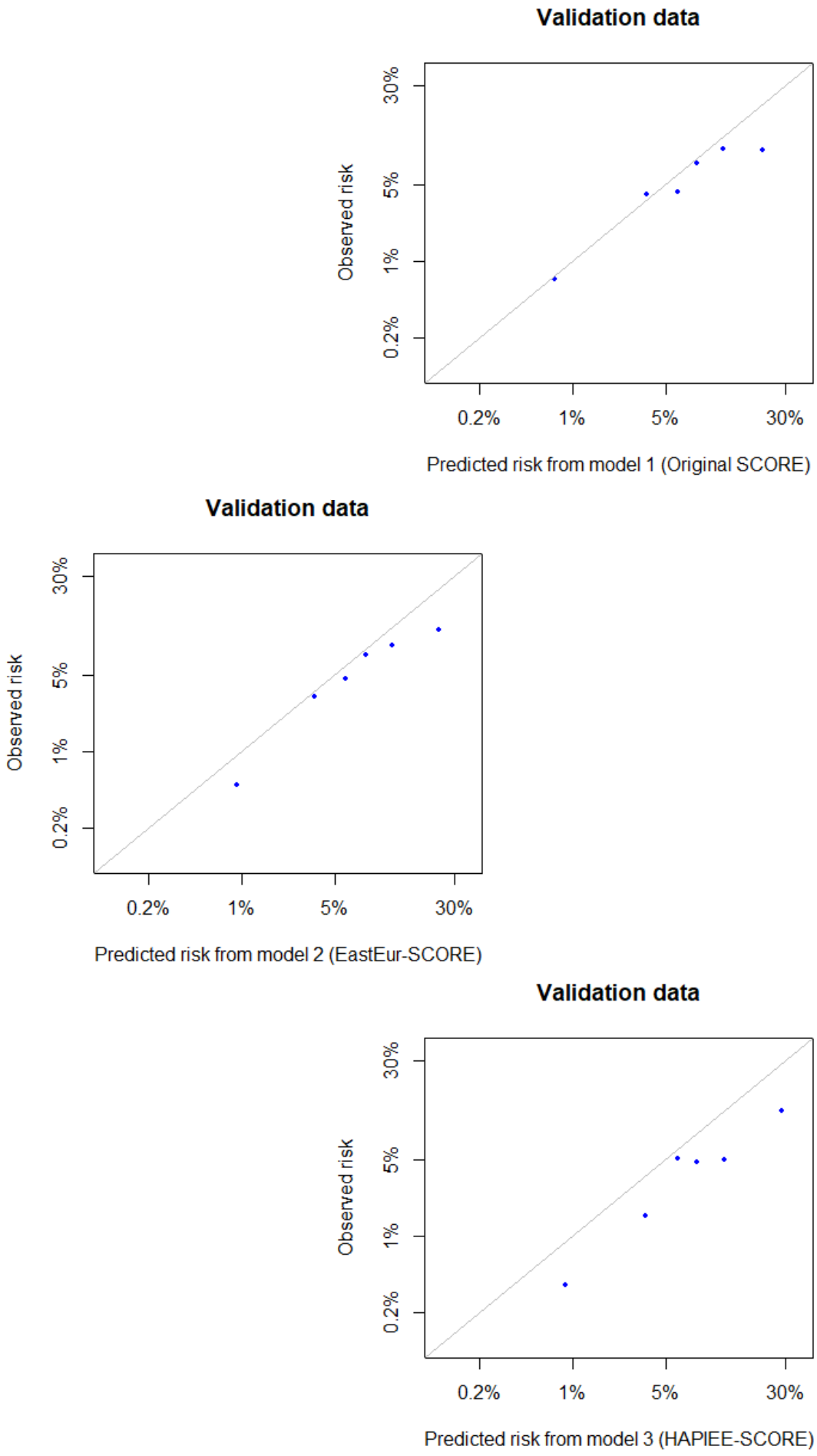


Figure 44. Calibration plots in the validation data.

3.3.3. Discrimination

The original SCORE model already had good discrimination performance in both derivation and validation datasets (C statistics of 0.78 and 0.83, respectively, table 13). This withstanding, the process of updating SCORE to create EastEur-SCORE resulted in impressive improvements to discrimination (ΔC = 0.04 and 0.02).

Furthermore, the process of adding new predictors to create HAPIEE-SCORE additionally led to substantial improvements (ΔC = 0.02 and 0.014).

Taken together, these two steps improved the original SCORE discrimination by an outstanding degree (ΔC = 0.06 and 0.03).

	Name of model		
	Original SCORE	***EastEur-SCORE***	***HAPIEE-SCORE***
Derivation data			
C-statistic	0.783 (0.735 - 0.831)	0.818 (0.774 - 0.862)	0.840 (0.800 - 0.880)
change in C-statistic	0.035	0.022	
			0.057
Validation data			
C-statistic	0.832 (0.769 - 0.896)	0.851 (0.791 - 0.910)	0.865 (0.806 - 0.923)
change in C-statistic	0.019	0.014	
			0.033

Table 13. Discrimination performance of three cardiovascular prediction models, as measured by Harrell's C-statistic (and 95% CI). *Numbers without 95% CI denote changes in C-statistic, when comparing two models.*

3.3.4. Classification

Table 14 shows classification performance across the three models. As previously described, the original SCORE overpredicted risk in the derivation set. This phenomenon can also be seen in higher sensitivity (77%) and lower specificity (64%) in the top row, when compared to the newly derived models.

In the validation sample, the newly created models also tended towards over prediction. This too is visible in the pattern of increasing sensitivity and decreasing specificity, when looking down the three validation rows.

Cohort	Model		predicted low risk	predicted high risk	TOTAL	Sensi-tivity	Speci-ficity	PPV	NPV
Derivation	SCORE	Whole sample	9,250	5,348	14,598	77%	64%	4.9%	99%
		Cases only	77	261	338				
		Controls only	9,173	5,087	14,260				
Derivation	EastEur-SCORE	Whole sample	11,376	3,222	14,598	66%	79%	7.0%	99%
		Cases only	114	224	338				
		Controls only	11262	2998	14,260				
Derivation	HAPIEE-SCORE	Whole sample	11,431	3,167	14,598	69%	79%	7.3%	99%
		Cases only	106	232	338				
		Controls only	11325	2935	14,260				
Validation	SCORE	Whole sample	3,814	818	4,632	58%	83%	6.5%	99%
		Cases only	38	53	91				
		Controls only	3776	765	4,541				
Validation	EastEur-SCORE	Whole sample	3,698	934	4,632	69%	81%	6.7%	99%
		Cases only	28	63	91				
		Controls only	3670	871	4,541				
Validation	HAPIEE-SCORE	Whole sample	3,396	1,236	4,632	81%	74%	6.0%	99%
		Cases only	17	74	91				
		Controls only	3379	1162	4,541				

Table 14. Classification performance of three cardiovascular prediction models.
PPV = Positive Predictive Value; NPV = Negative Predictive Value

3.3.4.1 Modelled clinical effect from statins

To better understand the trade-offs between these two competing tensions, a brief cost-effectiveness simulation is presented in Table 15. Looking at the very top row (performance of the original SCORE in the derivation set), we have previously seen how this instance showed greater over prediction than in any of the other models. In this table, the same pattern of over prediction means that the model incorrectly labels people of intermediate risk as being of high risk, thereby recommending for them to consider statins, which in turn increases the programme costs. Since many of these statins will be recommended unnecessarily, then this reduces the programme cost-effectiveness. Indeed, for the top row the ""€ / QALY" metric has the least favourable cost-effectiveness estimate of all the rows shown (€3,788, whereby large numbers are unfavourable. Of note, Western European countries generally tend to consider healthcare services that cost less than 20 000 or 30 000 per QALY to be sufficiently efficient to fund at scale).

Recalibrating SCORE in the derivation data led to much fewer people being treated with statins and hence a smaller net QALY gain than in the first row. However, given the substantially lower programme cost (€384,109 Vs. €637,560 to screen 10 000 people and treat those who are identified as being at high risk), the overall cost-effectiveness estimate improved by 35% to €2,475. Augmenting SCORE with the novel risk factors led to slightly fewer people being treated, as well as slight increases to net QALY gains made to overall health, thereby slightly improving cost-effectiveness by a further 6% to €2,329. However, it is unclear to what degree these estimates may be biased by overfitting the models used to construct these estimates.

Cohort	Model	Number Needed to Screen	Number Needed to Treat	Cost (to screen and treat 10'000)	QALY changes due to			Cost per QALY (efficiency)	Interpretation
					CVD mortality averted	Side effects caused	Net (effective-ness)		
Derivation	SCORE	243	85	€ 637,560	210	42	168	€ 3,788	Most inefficient, nearest to the *"treat everybody"* strategy (high cost and high output).
Derivation	EastEur-SCORE	283	58	€ 384,109	181	25	155	€ 2,475	
Derivation	HAPIEE-SCORE	274	55	€ 377,553	187	25	162	€ 2,329	Best model in derivation cohorts (as measured by all metrics).
Validation	SCORE	380	63	€ 306,232	134	20	114	€ 2,687	Most ineffective, nearest to the *"treat nobody"* strategy (low cost and low output).
Validation	EastEur-SCORE	320	60	€ 349,659	160	23	136	€ 2,563	Best efficiency (and good effectiveness).
Validation	HAPIEE-SCORE	272	68	€ 462,717	187	31	157	€ 2,951	Best effectiveness (and good efficiency).

Table 15. Clinical and economic models, applying six modelling scenarios to a hypothetical cohort of 10 000 people.

QALY = Quality Adjusted Life Year. CVD = Cardiovascular Disease.

Red font denotes quantities where small numbers are preferred. Red shade denotes undesirable performance.

Blue font denotes quantities where large numbers are preferred. Blue shade denotes desirable performance.

Net QALY change (a favourable outcome) is calculated as CVD mortality averted (favourable outcome) minus Side effects caused (unfavourable outcome).

In the validation dataset (which should be less prone to overfitting), the original SCORE seemed to slightly under predict risk, as programme costs (€306,232) and clinical benefits here were smaller than with the other five scenarios. The recalibrated SCORE increased costs slightly, but this was offset by much better net QALY gains as more people were prevented from developing CVD. This clinical benefit outweighed the detriment to programme costs, and so overall cost-effectiveness (€2563) increased by 5% when compared to the Original SCORE in the validation data. Finally, after adding the novel risk factors, the augmented SCORE tended towards slightly over predicting risk. This was not large enough to tip the balance of clinical effects towards a dominant increase in harm from side effects, but rather a dominant increase in benefit from preventing CVD, thereby leading to small improvements to net QALY gains. However, since the costs required to achieve this were considerable, augmentation with novel risk factors led to a 15% decrease in cost-effectiveness, when compared to the recalibrated model without novel risk factors.

These brief examples of modelling approximate cost-effectiveness scenarios illustrate the importance of selecting the optimal intervention threshold. Importantly, programme costs can vary by as much as twofold simply by using slightly different models (with different degrees of calibration or interventional thresholds) For example, the final HAPIEE-SCORE may well outperform the other models in the validation dataset, if the 5% threshold of clinical intervention was adjusted downwards to correct for this model's tendency to over predict risk. This can be thought of as "altering the threshold to fit the model". However, since I had specified a priori that I will only be interested in clinical changes at the 5% threshold, I am not going to optimize the threshold for the final model any further, and I will not perform any final recalibration of this model either.

This withstanding, reclassification plots, shown in the next section, enable a preliminary inspection of potential reclassification across a range of interventional thresholds.

3.3.4.2 Reclassification plots

Figure 45 illustrates reclassification in the derivation cohort, when comparing the original SCORE against the recalibrated model. If all dots lie along the diagonal, this denotes no change in risk prediction between the two models. Dots which lie upwards and left of the diagonal line have been reclassified upwards in risk, while dots which lie downwards and right of the diagonal have been reclassified downwards in risk with the new model. Of less importance are dots appearing in the bottom-left quadrant and dots appearing in the top-right quadrant, since all these dots denote various degrees of reclassification that is clinically of little consequence. In contrast, red dots appearing in the top left corner would suggest that the new model is clinically favourable (since it is better able to classify cases as being at high risk, who were previously classified as low risk). Along similar lines, blue dots appearing in the bottom right corner would suggest that the new model is clinically favourable (since it is better able to classify controls.

Looking at figure 45, both blue and red dots are slightly more common down-and-right of the diagonal. This is consistent with the interpretation that the original SCORE overpredicted risk, which has now been corrected in EastEur-SCORE (model 2) following recalibration. The pattern is particularly prominent among the blue dots (denoting controls). In addition, a small cluster of red dots is seen above the diagonal, in the top right corner. This suggests that true positives have correctly been classified upwards, consistent with the anticipated benefits from adding a new smoking parameter (ex/light smoker), as well as more sophisticated nonlinear modelling of the risk functions for cholesterol and blood pressure. However, from this plot it is unclear whether these changes are clinically meaningful, as both blue and red dots are found in the clinically meaningful quadrants (top-left; bottom-right).

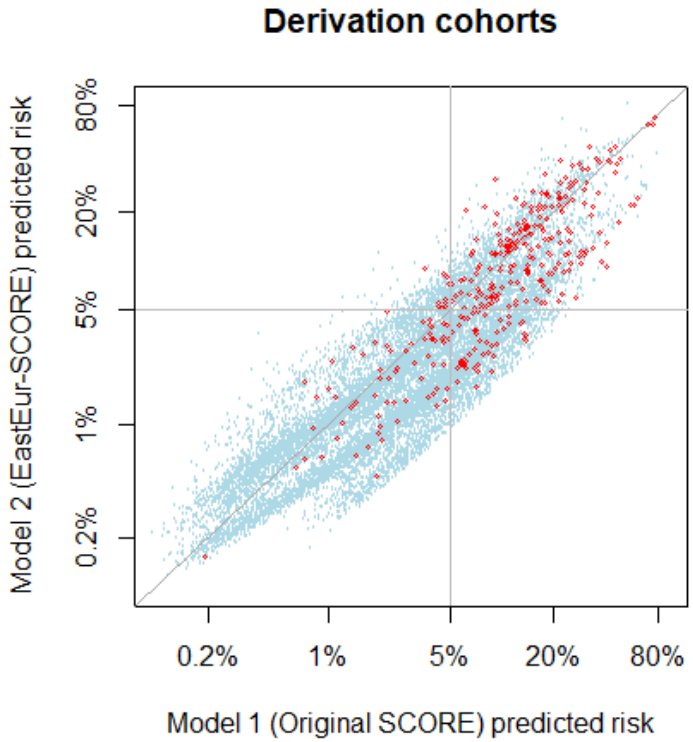


Figure 45. Reclassification in the derivation cohorts, comparing the original SCORE model (x-axis) against the recalibrated model (y-axis).

Red dots denote participants who died of CVD during the follow-up.
Blue dots denote participants who did not die of CVD during the follow-up.

Figure 46 illustrates the same reclassification, this time in the validation cohort. There are more data points with very low risk, possibly reflecting the slightly younger age profile and/or potential selection bias from a stronger “healthy volunteer” effect. In both the derivation and validation datasets, the curvature of the blue dots across the risk spectrum appears as a result of transitioning from a Weibull model to a Proportional Hazards model, as the hazard associated with age is modelled slightly differently across these two.

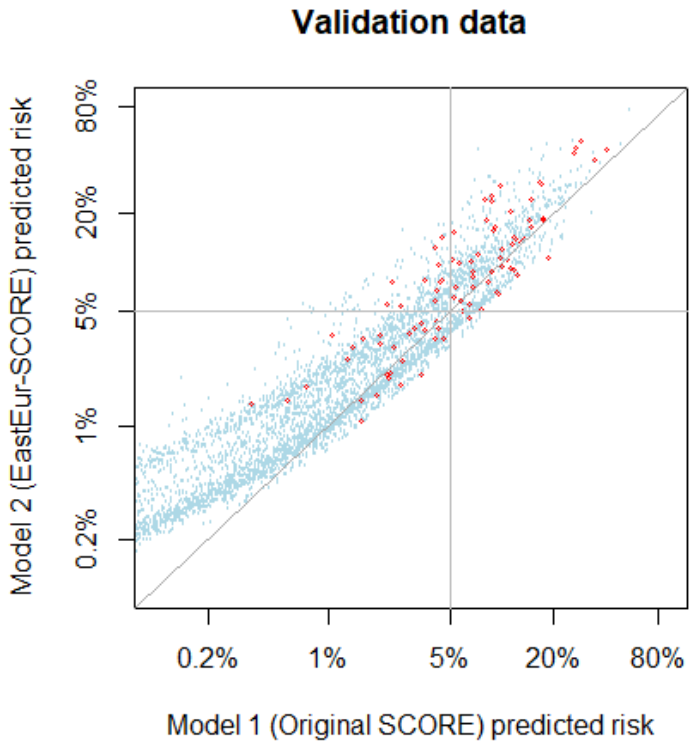


Figure 46. Reclassification in the validation cohort, comparing the original SCORE model (x-axis) against the recalibrated model (y-axis).
Red dots denote participants who died of CVD during the follow-up.
Blue dots denote participants who did not die of CVD during the follow-up.

In contrast to the derivation data (figure 45), the validation cohort in figure 46 suggest a slightly opposite effect: that there is a slight upclassificaion of risk, among both cases and controls. One outcome of this is a relatively large number of red dots in the top left quadrant. These are participants who previously were classified as “low risk” (and not eligible for interventions like statins), but who have now been correctly classified as “high risk”. This is consistent with the increase in sensitivity across the two models (from 58% to 69%). Fortunately, the corresponding decrease in specificity, on account of blue dots also moving into the top left quadrant, is not as pronounced (from 83% to 81%), suggesting that the reclassified model might be generally superior.

Figure 47 illustrates reclassification after the second step, where the recalibrated EastEur-SCORE was augmented with seven new risk factors. The top left corner contains more red dots than the bottom right corner, consistent with an interpretation of superior discriminatory power in the newly augmented model. In addition, I also modified the location of the horizontal and vertical lines, thereby changing the size of these two quadrants, to denote alternative clinical interventions thresholds that are near to 5%. These plots are not shown in the thesis, but my inspection suggested that superior discriminatory power may be present across a range of intervention thresholds, such as from 2% to 8%. If the threshold is raised above 8%, then those few individuals who do develop CVD tend to do so for unknown risk factors that are not captured by my final augmented model. If the threshold is raised below 2%, then those few individuals who do develop CVD tend to do so for unknown risk factors that are not captured by my final augmented model.

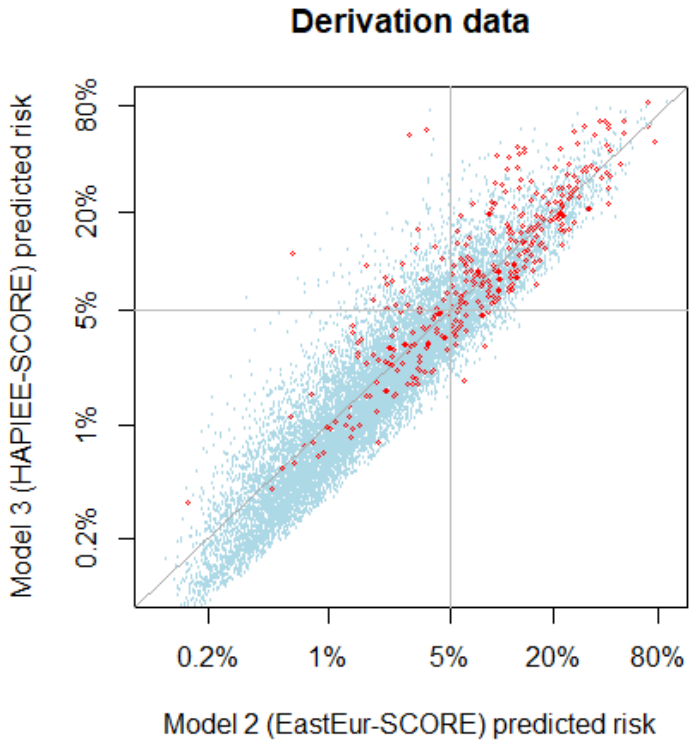


Figure 47. Reclassification in the derivation cohort, comparing the recalibrated model (x-axis) against the augmented model (y-axis).

Red dots denote participants who died of CVD during the follow-up.
Blue dots denote participants who did not die of CVD during the follow-up.

Figure 48 illustrates the same reclassification in the validation data. Again, the top left corner contains more red dots than the bottom right corner, consistent with an interpretation of superior discriminatory power in the new model. Again, modification of the 5% clinical interventions thresholds here suggested that superior discriminatory power may be present across virtually the entire range of intervention thresholds (from 2% to 50%).

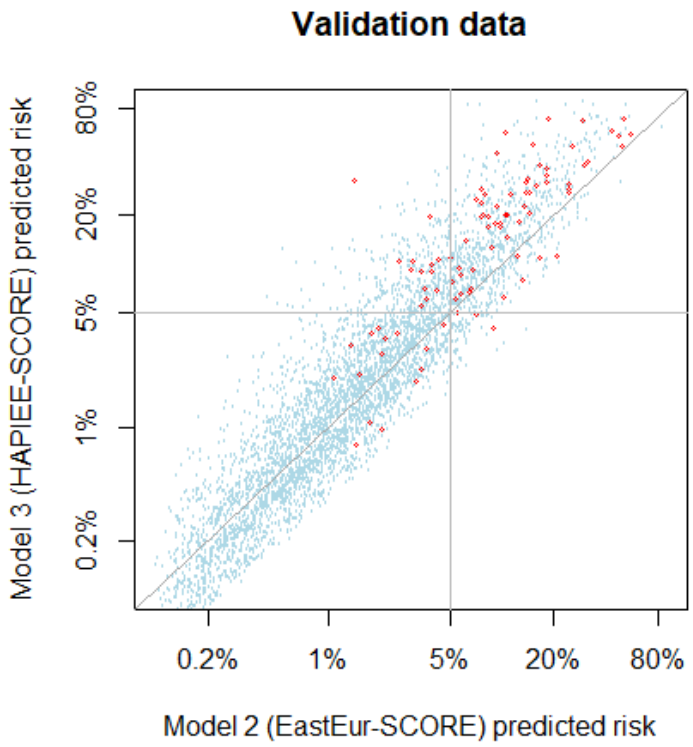


Figure 48. Reclassification in the validation cohort, comparing the recalibrated model (x-axis) against the augmented model (y-axis).
Red dots denote participants who died of CVD during the follow-up.
Blue dots denote participants who did not die of CVD during the follow-up.

3.3.4.3 Reclassification tables

The next section will describe 6 tables, where each table makes a detailed comparison of the performance of two risk prediction models. In this case, data are presented with exclusive interest in differences across the clinically meaningful risk thresholds (5% risk of CVD mortality in 10 years' time).

(The sole exception to this focus is how *Continuous NRI* is still reported at the bottom for comparison, since this measure has been commonly reported in the existing literature. Theoretically, *Continuous NRI* assumes that there is not only one risk threshold [as is assumed by *Binary NRI*] or a handful of risk thresholds [as is assumed by *Continuous NRI*] but that there are even more – a total of infinite risk thresholds. As the number of thresholds increases, then NRI itself tends to increase in number. However, it can be argued that Binary forms of NRI are clinically more informative and useful, than Continuous forms of NRI. This is detailed in Appendix 2. For this reason, my use of the term "NRI" or "reclassification" will always refer to "Binary NRI" unless otherwise specified.)

Data are further stratified to examine upwards and downwards movement across this 5% risk threshold for cases and controls.

Table 16 shows the outcomes for reclassification, when the derivation data are used to compare the original SCORE (model 1) against the recalibrated SCORE (model 2, a.k.a. EastEur-SCORE). When looking only among the cases (top table), then 3% of participants changed correctly from a previously incorrect "low risk" classification to now being correctly classified as "high risk". However, 14% of participants changed incorrectly from a high to a low risk classification. This resulted in a negative Binary NRI among cases (-11%). Among controls, the Binary NRI was positive (+15%). This pattern of change is consistent with the interpretation that the new model corrected for previous miscalibration. These two reclassification quantities can be added to yield a summative Binary NRI of +0.04. However, if 20x more weight is given to the changes

in classification among cases, when compared to changes in classification among controls (as is done by the Net Benefit metric), then Net Benefit is negative (-0.10). This is to be expected if the entire risk model 2 has been calibrated downwards in risk, as it has been in this case.

If considering risk reclassification at infinite risk thresholds, then the continuous NRI was positive (+0.19). However, I will place less value on this metric as I find it clinically less informative.

From model 1 (SCORE) to model 2 (EastEur-SCORE), derivation data

Among cases (observed events)

		New model (EastEur-SCORE)					
		predicted low risk	predicted high risk	TOTAL	Correctly upclassified	Incorrectly downclassified	NRI in cases
Old model (SCORE)	predicted low risk	67	10	77	3%		
	predicted high risk	47	214	261		14%	
	TOTAL	114	224	338			-11%

Among controls (observed no event)

		New model (EastEur-SCORE)					
		predicted low risk	predicted high risk	TOTAL	Incorrectly upclassified	Correctly downclassified	NRI in controls
Old model (SCORE)	predicted low risk	8885	288	9173	2%		
	predicted high risk	2377	2710	5087		17%	
	TOTAL	11262	2998	14260			15%

Summary

Net Reclassificaiton Improvement, continuous	0.19	(0.08 to 0.29)	P < 0.0001
Net Reclassification Improvement, binary/categorica	0.04	(-0.01 to 0.08)	P = 0.09
Net Benefit (with a casese:control weighting of 20:1)	- 0.10		

Table 16. Reclassification in the derivation cohorts, comparing the original SCORE against the recalibrated SCORE.
NRI=Net Reclassification Improvement.
Green highlight = large percentages here are desirable.
Red highlight = large percentages here are undesirable.

Table 17 shows subsequent reclassification performance, after adding seven new predictors (model 3) to the recalibrated model (model 2). There was a slight increase in reclassification performance among cases (+2%), alongside no change in reclassification among controls (+0.4%), altogether yielding a Binary NRI of 0.03. After up weighting the changes in performance among cases, there was a positive gain in Net Benefit (+0.02).

From model 2 (EastEur-SCORE) to model 3 (HAPIEE-SCORE), derivation data

Among cases (observed events)

		New model (HAPIEE-SCORE)					
		predicted low risk	predicted high risk	TOTAL	Correctly upclassified	Incorrectly downclassified	NRI in cases
Old model (EastEur-SCORE)	predicted low risk	90	24	114	7%		
	predicted high risk	16	208	224		5%	
	TOTAL	106	232	338			2%

Among controls (observed no event)

		New model (HAPIEE-SCORE)					
		predicted low risk	predicted high risk	TOTAL	Incorrectly upclassified	Correctly downclassified	NRI in controls
Old model (EastEur-SCORE)	predicted low risk	10671	591	11262	4%		
	predicted high risk	654	2344	2998		5%	
	TOTAL	11325	2935	14260			0.4%

Summary

Net Reclassificaiton Improvement, continuous	0.44	(0.33 to 0.55)	$P < 0.0001$
Net Reclassification Improvement, binary/categorical	0.03	(-0.01 to 0.07)	$P = 0.14$
Net Benefit (with a casese:control weighting of 20:1)	0.02		

Table 17. Reclassification in the derivation cohorts, comparing the recalibrated SCORE against the augmented SCORE.
NRI=Net Reclassification Improvement.
Green highlight = large percentages here are desirable.
Red highlight = large percentages here are undesirable.

Table 18 shows the total reclassification performance, after comparing model 1 against model 3. An overall downwards recalibration (possibly arising from the creation of model 2) can be seen in a negative NRI among cases (-9%). On the other hand, reclassification among controls was generally much improved (+15%), perhaps on account of the better discrimination created by model 3. Summing these two components in an unweighted manner yielded a Binary NRI of +0.07, but allocating appropriate clinical weighting led to a negative Net Benefit of -0.08. This may be expected in cases where the original SCORE overpredicted risk substantially, as the clinical Net Benefit measure is prone to rewarding over prediction since even slight increases in sensitivity are overshadowed (by 20 times) by any loss in specificity.

From model 1 (SCORE) to model 3 (HAPIEE-SCORE), derivation data

Among cases (observed events)

		New model (HAPIEE-SCORE)			Correctly upclassified	Incorrectly downclassified	NRI in cases
		predicted low risk	predicted high risk	TOTAL			
	predicted low risk	60	17	77	5%		
Old model	predicted high risk	46	215	261		14%	
(SCORE)	TOTAL	106	232	338			-9%

Among controls (observed no event)

		New model (HAPIEE-SCORE)			Incorrectly upclassified	Correctly downclassified	NRI in controls
		predicted low risk	predicted high risk	TOTAL			
	predicted low risk	8774	399	9173	3%		
Old model	predicted high risk	2551	2536	5087		18%	
(SCORE)	TOTAL	11325	2935	14260			15%

Summary

Net Reclassificaiton Improvement, continuous	0.45	(0.34 to 0.56)	P < 0.0001
Net Reclassification Improvement, binary/categorical	0.07	(0.02 to 0.11)	P = 0.005
Net Benefit (with a casese:control weighting of 20:1)	-0.08		

Table 18. Reclassification in the derivation cohorts, comparing the original SCORE against the augmented SCORE.
NRI=Net Reclassification Improvement.
Green highlight = large percentages here are desirable.
Red highlight = large percentages here are undesirable.

Table 19 shows reclassification in the validation cohort, comparing models 1 and 2. Here, because the original SCORE was reasonably well calibrated (perhaps just slightly under predicting risk but not much), the NRI was positive in cases (+11%) with minor loss of performance among controls (-2%). The unweighted Binary NRI was 0.09, and weighting resulted in a positive gain in Net Benefit (0.11).

From model 1 (SCORE) to model 2 (EastEur-SCORE), validation data

Among cases (observed events)

		New model (EastEur-SCORE)			Correctly upclassified	Incorrectly downclassified	NRI in cases
		predicted low risk	predicted high risk	TOTAL			
Old model (SCORE)	predicted low risk	27	11	38	12%		
	predicted high risk	1	52	53		1%	
	TOTAL	28	63	91			11%

Among controls (observed no event)

		New model (EastEur-SCORE)			Incorrectly upclassified	Correctly downclassified	NRI in controls
		predicted low risk	predicted high risk	TOTAL			
Old model (SCORE)	predicted low risk	3572	204	3776	4%		
	predicted high risk	98	667	765		2%	
	TOTAL	3670	871	4541			-2%

Summary

Net Reclassificaiton Improvement, continuous	- 0.09	(-0.28 to 0.10)	P = 0.34
Net Reclassification Improvement, binary/categorical	0.09	(0.02 to 0.16)	*P* = 0.02
Net Benefit (with a casese:control weighting of 20:1)	0.11		

Table 19. Reclassification in the validation cohorts, comparing the original SCORE against the recalibrated SCORE.
NRI=Net Reclassification Improvement.
Green highlight = large percentages here are desirable.
Red highlight = large percentages here are undesirable.

Table 20 shows the change in validation cohorts, across model 2 and 3. The changes are very similar to the previous changes from models 1 and 2: large NRI improvements seen among cases (+12%), coupled with minor deterioration among controls (-6%) led to an overall unweighted binary NRI of 0.06, and a large increase in weighted Net Benefit (0.12).

From model 2 (EastEur-SCORE) to model 3 (HAPIEE-SCORE), validation data

Among cases (observed events)

		New model (HAPIEE-SCORE)			Correctly upclassified	Incorrectly downclassified	NRI in cases
		predicted low risk	predicted high risk	TOTAL			
Old model (EastEur-SCORE)	predicted low risk	15	13	28	14%		
	predicted high risk	2	61	63		2%	
	TOTAL	17	74	91			12%

Among controls (observed no event)

		New model (HAPIEE-SCORE)			Incorrectly upclassified	Correctly downclassified	NRI in controls
		predicted low risk	predicted high risk	TOTAL			
Old model (EastEur-SCORE)	predicted low risk	3309	361	3670	8%		
	predicted high risk	70	801	871		2%	
	TOTAL	3379	1162	4541			-6%

Summary

Net Reclassificaiton Improvement, continuous	0.53	(0.37 to 0.68)	P < 0.0001
Net Reclassification Improvement, binary/categorical	0.06	(-0.02 to 0.14)	P = 0.16
Net Benefit (with a casese:control weighting of 20:1)	0.12		

Table 20. Reclassification in the validation cohorts, comparing the recalibrated SCORE against the augmented SCORE.
NRI=Net Reclassification Improvement.
Green highlight = large percentages here are desirable.
Red highlight = large percentages here are undesirable.

Altogether, table 21 illustrates the change from model 1 to 3 in the validation data. As expected, this reports larger magnitudes of the previously described pattern: large improvements to reclassification among cases (+23%), coupled with small deterioration among reclassification among control (-9%). These two combined to yield an unweighted Binary NRI of 0.14. If weighted, there was a very large increase in Net Benefit (0.23).

From model 1 (SCORE) to model 3 (HAPIEE-SCORE), validation data

Among cases (observed events)

		New model (HAPIEE-SCORE)					
		predicted low risk	predicted high risk	TOTAL	Correctly upclassified	Incorrectly downclassified	NRI in cases
Old model (SCORE)	predicted low risk	14	24	38	26%		
	predicted high risk	3	50	53		3%	
	TOTAL	17	74	91			23%

Among controls (observed no event)

		New model (HAPIEE-SCORE)					
		predicted low risk	predicted high risk	TOTAL	Incorrectly upclassified	Correctly downclassified	NRI in controls
Old model (SCORE)	predicted low risk	3310	504	3814	11%		
	predicted high risk	86	732	818		2%	
	TOTAL	3396	1236	4632			-9%

Summary

Net Reclassificaiton Improvement, continuous	0.13	(-0.03 to 0.28)	P = 0.12
Net Reclassification Improvement, binary/categorica	0.14	(0.04 to 0.25)	P = 0.006
Net Benefit (with a casese:control weighting of 20:1)	0.23		

Table 21. Reclassification in the derivation cohorts, comparing the original SCORE against the augmented SCORE.
NRI=Net Reclassification Improvement.
Green highlight = large percentages here are desirable.
Red highlight = large percentages here are undesirable.

3.3.5. Decision Curve Analysis

A decision curve analysis was used as a sensitivity check, to explore whether the benefits reported above are likely to hold in cases where the intervention threshold (where interventions are recommended to those whose predicted risk is above 5%) is relaxed so that this could be anything between 0-20%.

Figure 49 reports results from the derivation dataset. Since the red line is consistently above the dotted line, this suggests that the recalibrated SCORE provides a large clinical net benefit across a range of interventional thresholds. Next, comparing the red and green lines suggest that the incorporation of seven new predictors in the HAPIEE model provides a small degree of additional net benefit for most scenarios (except for cases where the intervention threshold is between 0.09 and 0.12 where the EastEur and HAPIEE models are indistinguishable).

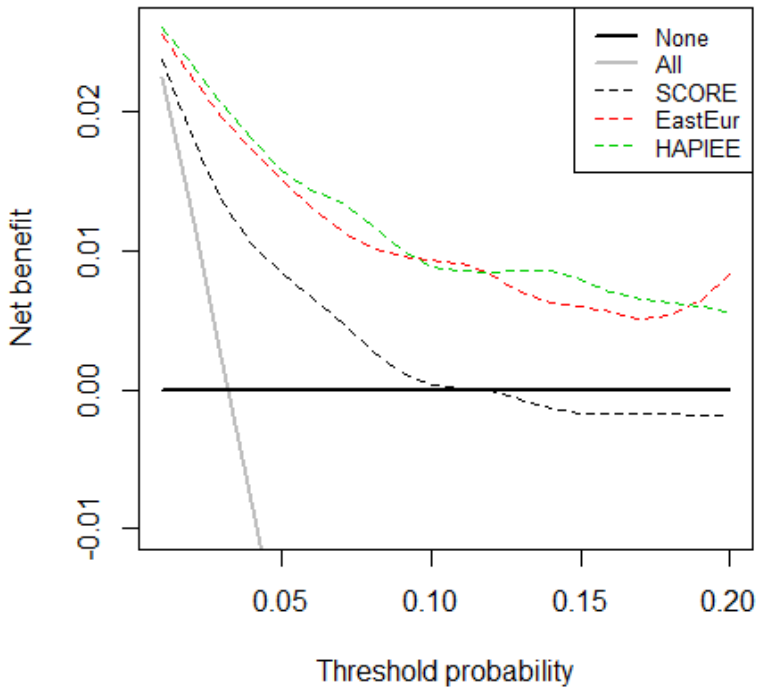


Figure 49. Decision Curve Analysis, in the derivation dataset. *X-axis denotes the threshold at which clinical interventions are recommended (a scenario that is more concerned with side effects and costs is left of 0.05, while a scenario that is more concerned with preventing CVD at any cost is right of 0.05). SCORE = The original SCORE model. EastEur = the recalibrated model 2. HAPIEE = the augmented model 3 that incorporates 7 additional predictors.*

Figure 50 reports the results from decision curve analysis performed on the validation dataset. All three models performed reasonably similarly across a range of interventional thresholds. This is consistent with the observation that the SCORE model performed relatively well in the validation data (when compared to the derivation data), and thus the baseline standard was already quite high. Inspection of the small changes in net benefit suggests that the final HAPIEE model appears to be superior in cases where the interventional threshold ranges from 2% to 6%.

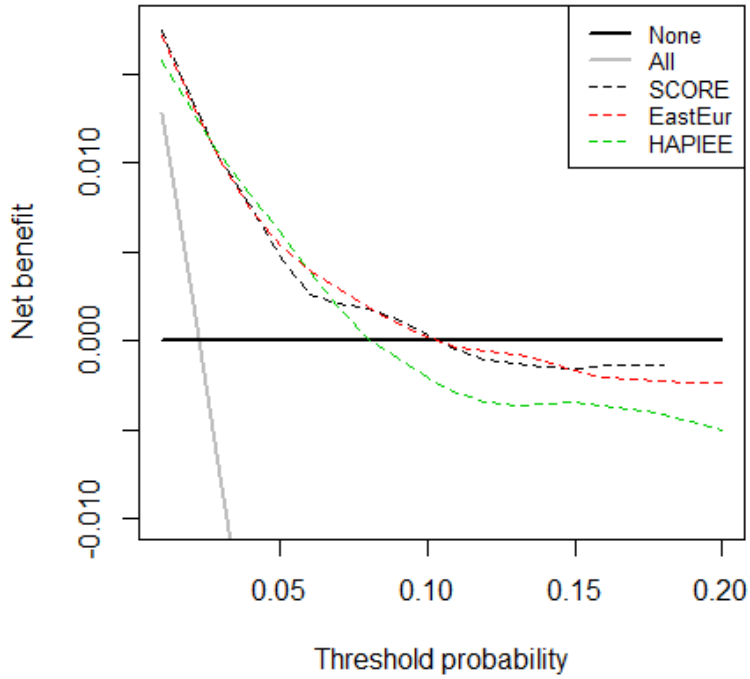


Figure 50. Decision Curve Analysis, in the validation dataset. *X-axis denotes the threshold at which clinical interventions are recommended (a scenario that is more concerned with side effects and costs is left of 0.05, while a scenario that is more concerned with preventing CVD at any cost is right of 0.05). SCORE = The original SCORE model. EastEur = the recalibrated model 2. HAPIEE = the augmented model 3 that incorporates 7 additional predictors.*

4. Discussion

4.1. Mendelian randomization

In this Mendelian randomisation study, I found strong genetic support for the hypothesis that longer education has a causal effect on lowering the risk of coronary heart disease. My findings using genetic data, which can be considered as 'nature's randomized trials',(257) were consistent with data from observational studies, and I found little evidence that these results may be driven by genetic pleiotropy. More specifically, 3.6 years of additional education (similar to an undergraduate university) degree is predicted to translate into about a one-third reduction in the risk of CHD.

4.1.1. Comparison with previous studies

There is a vast body of observational studies across a range of settings that show an association between education and CHD. In contrast, there have been comparatively few studies that have explicitly investigated the causality of this relationship. The existing studies come from three domains. First, analyses of natural experiments have compared mortality before and after changes to compulsory schooling laws. For example, by looking at mortality rates in countries before and after the introduction of national legislation that increased minimum education. In the Netherlands, such changes were associated with reductions in all-cause mortality.(121) In the UK, the largest study so far reported causal effects on improving physical activity, BMI, blood pressure, diabetes, CHD, and all-cause mortality.(122) An extension of this design is to compare geographical areas, such as the various states within the USA. These studies initially suggested a large effect on all-cause mortality, but this effect disappeared when state-specific baseline trends were taken into account.(123, 124) In Sweden, an intervention to extend compulsory schooling throughout a 13-year transition period in a stepped-wedge design across multiple municipalities reported lower all-cause mortality in those deaths occurring after age 40 (equivalent to HR of death=0.86 (0.77-0.96) per 3.6 years of additional education).(125)

Another source of causal inference comes from studies on monozygotic twins. Within each pair, both twins are exposed to the same set of genetic exposures (and also some environmental exposures, called the “shared environment”). Consequently, any difference in disease outcome between twins cannot arise from genetic effects. If differences in outcome associate with differential exposure to nonshared features of the environment (such as one twin pursuing education longer than the other twin), and if the magnitude of this association is comparable to that seen in the general population, then this makes less likely the possibility that the observational association is confounded by genetic (or shared environmental) factors. While the twin method does not eliminate the possibility of confounding from other factors in the nonshared environment, it is a design with which to eliminate the possibility of confounding from genetic factors. Twin studies conducted in Denmark initially found evidence both for and against causal effects from education to mortality and CHD incidence.(126, 258) The largest study to date from Sweden (which has twice the statistical power of the previous largest study) found strong evidence for causal effects.(130) There, the association between years of education and lifespan did not attenuate at all, when comparing the conventional population-based analysis against the between-twin analysis. Hence the twin literature suggests that, while there are only a handful of sufficiently powered studies, shared environmental factors (such as parenting) are less likely to cause substantial confounding. It also suggests that confounding from genetic factors (such as genetic differences in drive, motivation, personality or innate intellect; all of which may predispose towards longer education) might not account for the observational associations between education and disease. A parallel domain of research, using data from millions of non-identical siblings (that sometimes reached 100-times larger sample sizes than the twin studies), has also observed little attenuation of the association between education and subsequent mortality, when comparing the general population analysis with the within-sibling analysis.(131, 132) As with twin studies, this also suggests that

environmental and genetic factors shared by the siblings, are unlikely to confound the observational association seen between education and disease. While twin and sibling studies both leave open the possibility of confounding from nonshared environmental factors, taken together with these results (using an entirely different method), the wider body of evidence is more compatible with a causal interpretation, suggesting that changes in education may lead to a reduction in CHD.

Finally, some recent studies have also looked at specific genetic variants for education. An association was found between parental longevity and genetic markers for education in their offspring.(134) However, causal directions and pleiotropy were not tested in this study. Others have used conventional Mendelian randomisation and found genetic variants for education to predict myopia and dementia.(136) However, these studies did not investigate pleiotropy of their genetic instruments. No Mendelian randomisation studies of socioeconomic exposures have investigated any other disease outcome, such as cardiovascular diseases. Furthermore, most of the other designs listed above (including natural experiments, twin and sibling designs) have reported outcomes for all-cause mortality. Few have reported cardiovascular mortality and virtually none have reported fatal/nonfatal CHD, as I did.

4.1.2. Strengths and limitations

My study has important strengths. I investigated the causality of the association between an easily measured socioeconomic factor (i.e. education) and a common disease (i.e. coronary heart disease). I applied the Mendelian randomisation design, which in conjunction with findings from other study designs, should improve our understanding of causality by reducing bias from confounding. By integrating summary-level data from over half a million individuals, my study was well-powered to derive robust causal effect estimates, and also powered for multiple sensitivity analyses (which typically require larger sample sizes). I used recent state-of-the-art methodological developments to thoroughly explore the possibility of pleiotropy in the genetic variants analysed, for which I found little evidence.

My study also has some limitations. First, it is possible that the genetic variants associated with education may instead mark more generic biological pathways (such as vascular supply or mitochondrial function), which could enhance systemic fitness, thereby leading to parallel increases in cognitive and cardiac function.(23, 129) In this scenario, which could violate the InSIDE assumption, policy interventions to raise education may not translate into lower heart disease incidence.

However, such a scenario is less likely to lead to the consistent set of results I found across my sensitivity analyses, as this would require that pleiotropy occurs in a rather contrived scenario where: i) the InSIDE assumption is violated (so that MR-Egger is biased); ii) least 50% of the information comes from SNPs with highly pleiotropic effects on heart disease; iii) and these pleiotropic effects occurred in such a way as to make the causal estimates on heart disease appear very similar to one another.

At present, there are no definitive tests with which to verify such assumptions, meaning that triangulating data from other sources and subjective judgement are needed to evaluate the plausibility of gross pleiotropic bias.(259) I believe such pleiotropy to be less likely for four reasons. First, the effects from genetic pleiotropy would have to coincide with the non-genetic associations observed in studies of monozygotic twins, and second they would also have to coincide with the non-genetic associations observed in natural experiments. Third, if education and CHD share some of their underlying genome-wide genetic architecture (as seen in the LD score regression), and if most of the top hits for education are strongly pleiotropic for CHD, then one might imagine the top hits for CHD to also pick up some of these pleiotropic traits. However, my reverse-direction Mendelian randomisation found a null estimate. Fourth, despite gaps in our understanding of the biological mechanisms through which these 162 SNPs influence education, they are disproportionately found in genomic regions that regulate brain development, they are enriched for biological pathways involved in neural development, and they are preferentially expressed in neural tissue.(234) Since these 162 SNPs do not appear to have any expression or enrichment in cardiovascular tissues, this further narrows the scope of pleiotropy: any potential pleiotropy might have to exert a large effect on CHD via predominantly neurological pathways (for example behaviours associated with obesity), as opposed to via global or systemic measures of fitness (such as mitochondrial function). Therefore on balance, I feel that the scenario where gross pleiotropy invalidates my sensitivity analysis is less consistent with the broader body of evidence, in comparison to the scenario where my sensitivity analyses are valid. If my sensitivity analyses are indeed valid, then policy interventions that mirror prolonged exposure to educational (as indexed by the genetic instruments) would, on balance, probably prevent heart disease.

As a second limitation, it is possible that these 162 SNPs are merely markers of cognitive ability and intelligence, which is the underlying factor that causes both higher education and lower heart disease. The body of evidence appears to suggest that intelligence can be a causal mediator along the pathway from education to heart disease.(260) This is most compatible with a simple “linear chain” model whereby the 162 SNPs I studied cause intelligence, which in turn predisposes towards longer education, which in turn increases intelligence further still, as a final result of which heart disease is ultimately prevented. This I believe to be the most likely causal pathway.

Alternatively, critics could posit that genetic predisposition to greater intelligence may theoretically be a *pleiotropic confounder*, which invalidates my MR and observational analyses. However, this model is difficult to reconcile with the causal effects measured by natural experiments and twin studies. In both of those cases, causal effects are elicited by varying the environmental (i.e. non-genetic) component of education. Although it is possible for other biases in these study designs to coincidentally align with this bias in my MR and observational results, the probability of this appears to be low.

As a third limitation, in order to arrive at policy recommendations, one would have to assume that *genetic predisposition towards higher educational attainment causes the same behavioural and physiological consequences as environmentally-acquired changes to educational attainment*, such as from a policy intervention. It may be however, that a year of additional education from genetic causes could trigger a different set of biological and behavioural mechanisms when compared to a year of additional education from policy change. Presently, we know very little about the mechanisms of these genetic effects. In the analyses I conducted in this study, I found some initial evidence that some of these genetic effects may be mediated via common cardiovascular factors like smoking, BMI and lipids. In keeping with this, policy changes to education in the USA and UK have also estimated some causal effects on smoking,

BMI, blood pressure and diabetes (122, 261), which are broadly consistent with my findings. Few studies have measured the causal effects of policy interventions on lipids. While a randomized controlled trial of education is difficult for CHD outcomes, due to approximately 50 years of lag, it may be conceivable for future research to measure effects on potential mediators, as these occur much sooner. A second response to this overall limitation is the analogy to other exposures (such as LDL-cholesterol and systolic blood pressure), where genetic effects have mirrored findings from environmentally-acquired changes (such as from randomized controlled trials of pharmacological therapies.(262, 263)). Taken together, while my study makes no direct inference on what health effects may stem from a policy intervention that successfully raises education, I am cautiously optimistic that a policy which successfully increases education should lead to reductions in heart disease.

As a fourth limitation, I assumed the absence of *dynastic effects*, an assumption that is broken when parental genes associate with parental behaviours that directly cause a health outcome in the child.(264) For example, parents with a genetic predisposition towards higher education may choose to feed their children a better diet. However, parental educational attainment has been shown to be a poor predictor of conventional cardiovascular risk factors in children.(265)

Fifth, the observational and genetic data originate predominantly from European origin samples in high-income countries. I am thus unable to generalize these estimates to other populations, particularly to low-income countries where cardiovascular diseases are less common. However, it may well be expected that socioeconomic factors mirror the pattern seen in other cardiovascular risk factors, where similar effects are typically seen across the world.(202) Along similar lines, my findings are predominantly valid to describe phenomena which may occur among the types of people who are likely to participate in scientific studies. It is unclear to what extent they will generalize to people who are typical non-responders.

Sixth, I do not know whether increasing education for those of least education will be as cardioprotective as increasing education for those with above-average education. Nonetheless, the linear relationship seen in the observational data is more compatible with an explanation of *dose-response* across the broad educational gradient.

Seventh, education is to some extent a social construct, whose attainment in one place and time may be influenced by different factors, when compared to attainment in another place and time. Indeed, the original GWAS study where these 162 SNPs were discovered analysed participants whose year of birth ranged from 1893 to 1989 – nearly a whole century. This challenges the generalizability and validity of my findings. However, the GWAS study also examined whether the polygenic score was more or less strongly associated with educational outcomes across the period 1930 to 1955 in Sweden, a time of considerable educational reform. There was no evidence of genetic effects operating differently across this period, suggesting that international differences in culture may be a greater source of heterogeneity. However, this GWAS study was conducted exclusively on European and North American settings, with broadly similar cultures. Hence although theoretically possible, I find it less of a concern for period or cultural effects to invalidate my analyses.

4.1.3. Potential mechanisms

4.1.3.1. Mediation via conventional cardiovascular risk factors

Previous work on the mechanism that might mediate the association between education and CHD have accounted for some (but not all) of this association. Traditional observational associations have typically estimated that the association between education and CHD attenuates by around 30-45% after statistical adjustment for health behaviours and conventional cardiovascular risk factors (including smoking, blood pressure and cholesterol). It has been reported that by measuring these mediators at two time points and modelling these as time-dependent covariates, the explained fraction attributable can increase to nearly 100% for outcomes such as all-cause mortality. However, one study which applied such a correction found that the fraction mediated for CVD mortality increased only from 29% to 45%.(266) Hence it is plausible for some of the mechanism from education to CVD to be mediated along additional pathways that are independent of the conventional CVD risk factors. This is consistent with the data from my observational analyses (figure 40 in section 3.2.3.), where I also found conventional risk factors to account for 42% of the association between education and CHD. Altogether, this suggests that conventional factors could account for around half of the association between education and CHD.(8, 266)

My Mendelian randomization study found genetic predisposition towards longer education to associate with improved smoking, BMI and blood lipid profiles (with some borderline results for blood pressure and risk of diabetes). In my analyses, I was unable to quantify to what degree these conventional risk factors (if indeed causal) could account for the pathway from education to CHD. This could be done in the future, for example by applying two-step Mendelian randomisation.(267, 268) I have interpreted my findings to mean that policy changes which increase educational attainment would probably reduce smoking and BMI among those who

have been pushed towards attaining higher education, when compared to a scenario without such policy change.

On the one hand, this interpretation may be incorrect if indeed dynastic effects (discussed above in section 4.1.2, under the fourth limitation) have confounded my causal analyses. For example, parents with genes predisposing towards greater education may create early-life conditions for their child that are more conductive towards heart health in ways that are completely independent of cognitive development and subsequent educational attainment (e.g. healthier diets or fewer adverse childhood experiences). Under this scenario, the fact that these children later undertake longer education may be irrelevant, in terms of influencing smoking and BMI outcomes. Indeed, socioeconomic gradients in BMI and smoking rates are already visible before age 16, which is consistent with a non-causal interpretation from own education to smoking. One analysis of causal effects as a result of raising the school leaving age in the UK showed mixed findings, with no causal effect found from education to smoking, but causal effects found from education to BMI.(122)

Given this mixed state of evidence, more work on the causal pathways operating in the first 25 years of life is warranted, perhaps even with randomized trial designs where possible. Until then, it appears sensible to suggest that interventions which seek to improve the early-life experiences of young children (both by increasing cognitive abilities as well as health behaviours) are likely to be of benefit, one way or another.

4.1.3.2. Mediation via other pathways

It is plausible that conventional cardiovascular risk factors may not completely account for the mechanism between education and CHD. Additional mechanistic hypotheses for investigations are required. First, spending more time in education could lead directly to greater health knowledge (from being in contact with staff or students with better health knowledge). Education could also increase health knowledge indirectly, by equipping people with generic cognitive and reading skills with which to make sense of health-related information (particularly online content, which is can be written by more educated authors). Better health knowledge could then lead to greater uptake of healthcare services. Alternatively, better health knowledge could then lead to better dietary behaviours. These putative pathways are beyond the scope of this thesis, but are promising avenues for further enquiry.

Second, education could improve people's job prospects, income, material conditions, social ranking, frequency of social contacts, marital status, and well-being. Socioeconomic and psychosocial –factors like these have previously been associated with education as well as CHD. Indeed my own observational mediation analyses suggested that altogether, they may account for an additional 35% of the pathway from education to CHD. This quantity should be interpreted very approximately, on account of large collinearity between these variables, residual measurement error, as well as any omitted confounders between mediator and outcome, all of which could bias quantity mediated up or down.(1)

Many of the putative mediators discussed above (including conventional cardiovascular risk factors, diet, uptake of healthcare, income, frequency of social contact, and overall well-being) are potentially amenable to clinical and/or public health interventions. These all present future opportunities to potentially lower health inequalities. Such attempts would benefit from rigorous evaluation where possible, in order to expand the comparatively light existing literature, on effective ways of reducing health inequalities.

4.1.4. What this study adds

Following exposure to a socioeconomic factor, there is often a long incubation period before the occurrence of common diseases (in this example, around 50 years). Consequently, this line of research is not particularly amenable to randomized controlled trials, which would otherwise settle questions of causality. This does not mean these associations are less worthy of investigation, particularly as large point estimates open up the possibility of potentially large public health gains. The solution is to triangulate evidence from multiple study designs, each with their own strengths and weaknesses. The limited studies to date have suggested that a causal effect is more likely than not to exist, between socioeconomic exposures and all-cause mortality. My study adds to this evidence by using an entirely new technique, which also suggests that a causal effect is more likely than not to exist between education and CHD.

4.1.5. Implications for researchers, clinicians and policymakers

The main question for future research is *"What mechanisms account for the strong association seen between genetic predisposition towards longer education, and substantially lower risk of CHD?"* Were it found that a health behaviour (such as diet) is an important mediator, then interventions on diet could become the cornerstone of policies designed to reduce health inequalities.

More molecular research is needed to delineate the mechanism, pleiotropic or not, through which these 162 education SNPs associate with cardiac outcomes. This could elucidate new aetiological mechanisms for CHD which, in turn, could lead to insights for potential drug discovery.

Although there remains uncertainty around the precise function of each of the 162 SNPs, their degree of pleiotropy with cardiac traits, and the mechanisms by which these genetic variants exert their cardioprotective

influence, policy conclusions can still be drawn from the current body of evidence. First, policies that raise education probably lead to non-health benefits, such as increased economic productivity, voter turnout, better governance, and improved life satisfaction.(269, 270) Second, there is very little evidence that raising education might subsequently harm health or wellbeing. Third, although rigorous scientific debate needs to continue on the health consequences of raising education, the current balance of opinion appears to weigh towards the side where raising education will probably improve a range of health outcomes (either to a smaller or larger degree). There has been little discussion about how to raise education in a manner that is practical, acceptable, affordable and sustainable. While my data makes no claims on this, I note that interventions should be accompanied by careful monitoring for unforeseen side effects, especially of those individuals who may not thrive when forced into extended educational settings, which may otherwise aggravate health inequalities. To briefly begin this discussion, one can imagine a range of policies by analogy to how clinicians, public health practitioners and policymakers encourage patients to stop smoking: by raising awareness (e.g. mass-marketing campaigns; personalized letters; individual counselling); convenience of access (e.g. changing the geographical dispersion of educational establishments; opportunities for flexible education); and/or finance (e.g. tuition fees; accommodation costs; stipends). One can also consider complementing some of these population-level policies with individual-level interventions (e.g. advising adolescents on whether to pursue higher education).

4.1.6. Conclusion

This Mendelian randomisation analyses found genetic support for the hypothesis that longer education plays a causal role in lowering the risk of coronary heart disease. Although it is difficult to completely rule out possible pleiotropic effects, the sensitivity tests available to us gave little evidence that these could have driven my findings. In conjunction with the results from other study designs, increasing education is likely to lead to large health benefits.

4.2. Mediation & international differences

I intended to investigate to what degree the pathways of various psychosocial and socioeconomic predictors of CVD may overlap with one another. In this large prospective cohort study, I found independent associations between six psychosocial factors and subsequent cardiovascular mortality following full mutual adjustment. Depression and social support factors did not substantially attenuate socioeconomic gradients in CVD mortality, suggesting that these mechanisms are more likely to be complementary, as opposed to being mediatory.

As a second objective, I investigated whether the previously unexplained difference in CVD rates between Russia and Central European countries could be attributable to differences in psychosocial and socioeconomic risk factors. Contrary to my hypothesis, the substantial international difference in baseline mortality was not reduced following adjustment for conventional, psychosocial or socioeconomic factors.

4.2.1. Strengths and limitations

Several limitations of this study should be considered when interpreting the results. First, these urban population samples are not necessarily representative of whole countries, as both exposures and mortality might be different in rural settings. Indeed, it is often the case for baseline mortality rates to be higher in rural settings, in which case the baseline rates I report may have been under-estimated. However, it is plausible for the international gradient that I describe to operate similarly across urban and rural settings. It is also plausible for biological and possibly psychosocial risk factors to operate similarly across urban and rural settings. Particularly in the absence of compelling evidence to support these putative interactive effects, it seems plausible to assume that the patterns I describe may generalize for most people, in each country under study.

Second, study participants may have been healthier than non-responders, making us blind to what happens among those facing the greatest health and social problems. This would have led to underestimated hazard ratios and population-attributable risk fractions. That said, our sample still detected considerable variation in cardiovascular mortality by country and socioeconomic status, as well as a considerable burden of psychological distress (e.g. 22% of participants screening positive for possible depression). Therefore it is possible that my analyses are biased towards the null, and the true prevalence and burden of CVD from psychosocial and socioeconomic risk factors may be even larger.

Third, although I used predominantly well-established instruments to assess psychosocial exposures, their self-reported nature may be affected by response bias. For example, those who report adverse profiles might be more neurotic or less conscientious in their personality, which might instead be the underlying cause of the observed associations. Indeed, both of these traits have been associated with CVD,(35) while their possibility of associating with the six psychosocial traits I describe has

been less studied. Reassuringly, in models where I additionally adjusted for the psychological trait “perceived control”, this did not substantially alter the other associations nor the summative conclusions (data not shown), suggesting that perhaps personality is not a particularly strong putative confounder.

Fourth, measurement error in mediators leads to underestimation in the degree of attenuation,(20) and more importantly, any unmeasured confounders between mediators and outcome could have biased the results of my mediation analyses in either direction.(19) As such, the mediation analyses I present should only be considered as very approximate and indicative only. Two-step Mendelian randomization is one potential future avenue, with which to investigate the direct causality towards mediators, and the amount that they causally mediate.

Fifth, this particular study makes no claims about causality due to its observational design. Some of the observational associations between CVD and social/psychosocial factors may be due to reverse causation or unmeasured confounding.(141) Further causal evaluation would benefit from analysis of policy interventions or Mendelian randomization methods applied to the other five risk factors that I describe.(224) Sixth, it is uncertain how generalizable these results might be to other countries, given the social history and high mortality rates in this region. Although such complex simultaneous analyses have not been previously performed in Western settings, it is nonetheless reassuring how most of the effect sizes I estimated in simple univariate models (adjusted for age and sex only) reflected broadly the effect sizes previously reported in Western literature. This suggests that perhaps these phenomena are similar across Western and Eastern Europe. This does not, however, allow any further extrapolation into potential effects in more diverse cultures, such as parts of Asia or Africa, that have received much less study.

Sixth, although the data I analysed offered no support for the hypothesis that the gap in mortality between Russia and Central Europe is

attributable to psychosocial or socioeconomic factors, I am unable to firmly reject this hypothesis. For example, if data collection was conducted with large measurement error, then these variables may simply have been too insensitive to detect the latent true causes of international variation. However even in this scenario, one would expect to have still seen at least a small narrowing of the international gap. In my analysis, the international gap did not narrow at all but widened after adjustment for psychosocial and socioeconomic factors.

Along similar lines, it is equally plausible that other psychosocial traits exist out there, which are the true causes of this international gap. However, these traits should not covary with any of the traits that were already measured by the HAPIEE study. As the HAPIEE study cast a relatively broad net over most of the known psychosocial and socioeconomic factors, this limits the state space where such undiscovered psychosocial or socioeconomic factors may potentially exist.

Seventh, it is quite possible that powerful psychosocial and/or socioeconomic effects may be operating at the level of the culture of the entire country, which could influence the way in which participants from one country interpreted the questions asked of them, as well as their perception of the culturally most appropriate response. For example, if all Russian residents had experienced a collective feeling of shame, guilt or embarrassment over the fate of their country (either for 1991-2008; or for throughout 1964-2008), and if this could trigger an over-compensatory mechanism, whereby all Russians would report their psychosocial/socioeconomic state with an optimistic bias (to project that the glass remains half-full). This could have resulted in in differential measurement error and systematic bias, by underreporting the true burden of psychosocial and/or socioeconomic risk factors seen in Russia. However, if this were the case, then one would imagine that the magnitude of association, between a given psychosocial risk factor and CHD, to be greater in Russia than Central Europe (since those people who subjectively reported mild psychosocial distress would probably have

experienced more severe psychosocial distress, from an objective perspective). However, such a pattern was not found in my empirical screen of interactive effects (available on https://doi.org/10.1371/journal.pmed.1002459.s008). This means that if indeed important and collective psychosocial factors remain to be discovered, then these should not covary with the variables that have already been collected by the HAPIEE study. Such a psychosocial variable should not vary much within one country, and should not predict CVD within a country. This again limits the state space where these undiscovered traits may potentially exist, making their presence increasingly unlikely.

A further extension of this idea is to posit that *all* Russians would be subject to an additional "Russian psychosocial stress" variable, which exerts a uniform CVD hazard, regardless of low or high SES status. While theoretically plausible and indeed quite possible, this is empirically very hard to test, as it is difficult to identify a suitable control or counterfactual group for comparison.

As well as limitations, this study also has important strengths. First, this is possibly the first prospective cohort study of psychosocial factors with a standardized methodology across multiple countries, where one country has twice the mortality rate of the others. Second, it is one of the largest prospective cohort studies that has assessed multiple psychosocial factors whilst concurrently controlling for all the conventional cardiovascular risk factors (including cholesterol and blood pressure). Third, most of the risk factors and covariates were measured using well-established and widely used psychosocial questionnaires and laboratory techniques.

4.2.2. Comparison with previous studies

4.2.2.1. Individual risk factors

Most of the associations I report are broadly consistent with prior studies, primarily from more affluent countries where all-cause mortality is often reported more commonly than my outcome of cardiovascular mortality.[2 4 5] However, I found current unemployment to be associated with an unusually high level of risk in my study (HR=2.96 [1.97-4.46] for all-cause mortality, age-sex adjusted), which is more than twice the estimate from a recent meta-analysis (HR=1.59 [1.42-1.77]).[3] This strong effect may be plausible, however, if unemployment protections are weaker in Eastern Europe than elsewhere, which might highlight policy weaknesses for intervention.

Literature on material conditions has mostly focused on area-level measures of exposure, while available studies at the individual level have often measured just 1-2 possessions,(22, 271) not aggregated such possessions into a summary score, or not controlled for blood pressure/cholesterol.(88) One comprehensive study in Russia did not find an association between material goods/amenities and all-cause mortality, once education was controlled for.(17) Our larger study found the opposite pattern: that *material amenities* was the principal socioeconomic factor, which displaced *education* in multivariate analysis. Our analysis is consistent with prior reports of how *education* and *material conditions* may be measuring the same underlying socioeconomic construct, and it emphasizes how *material amenities* might be a more sensitive socioeconomic predictor of cardiovascular mortality than *education*. Such phenomena, when two correlated measures “compete” for primacy in fully-adjusted models, may be prone to slight differences in the measurement error in one variable, as well as its statistical form (i.e. continuous or categorical). Furthermore, it is plausible for one risk factor to dominate in one cultural context, while another risk factor dominates in another cultural context.

4.2.2.2. Mediators of socioeconomic risk

Previous studies have suggested that *conventional* risk factors might account for around 30-50% of the primary association between socioeconomic factors and mortality,(11, 272, 273) consistent with my analysis of *education*. However, the proportion explained by conventional factors was much smaller for social support factors, such as only 3% for *lack of contact with friends*, and 8% for *single marital status*. Therefore, the role of conventional risk factors might be even smaller among the psychological factors such as social support and depression, factors that have been less studied to date.

This might be the largest study to look at whether primary socioeconomic gradients in CVD can be attenuated following adjustment with psychological factors such as depression and social support. In Whitehall II, the occupational gradient in non-fatal CHD did not attenuate substantially following inclusion of social support measures,(274) consistent with my findings. Instead, work-related stressors accounted for half of the occupational gradient in Whitehall II. I did not evaluate psychosocial work factors, as a recent meta-analysis has shown their association with CHD to be comparatively small.(275) Other studies have typically used single-item instruments,(276) or not assessed social support and conventional risk factors.(272, 273) My study has confirmed relatively robustly that depression and social support do not account for much of socioeconomic gradients, even in a large cohort with high prevalence of exposure and outcome.

4.2.2.3. Mediators of psychosocial risk

Our results suggest that the primary association between *depression/social support* and cardiovascular disease is unlikely to be mediated or confounded by conventional and other psychosocial risk factors, an area of relatively limited prior study.(277) For example, exposure to one of the three dimensions of low social support was not offset by excess in another dimension, suggesting that cardiovascular health may be protected by some contact with friends, family *and* a partner. It appears that, in contrast to socioeconomic factors, each psychological factor is distinctly separate, and does not relate to a common underlying construct. While attempts have been made to identify biomarkers of socioeconomic gradients (e.g. inflammation), comparatively few have attempted to discover the mediators of psychological factors.(167) My analysis suggests that depression is unlikely to be a major mediator of social support and socioeconomic pathways.

There is a paucity of studies investigating the potential mechanisms and mediators of psychosocial risk factors. I estimate, relatively crudely, that about one quarter of such hazards may be mediated by conventional CVD risk factors, and up to another quarter may come from other psychological factors such as depression. Measurement error in mediators biases estimates of mediation towards toward the null, so future studies with time-varying mediators may be able to demonstrate larger proportions mediated.(272) In case that such analyses fail to account entirely for the mechanistic pathways from psychosocial factors to CVD, then additional hypotheses for investigation are warranted. It may be particularly useful if future work could stratify analyses by gender. I will now present my initial suggestions, by explicitly using gender to denote putative pathways of greatest hypothesized magnitude.

First, it is plausible for socially isolated men (across all SES categories) to wait a longer time from symptom onset, before seeking formal or informal healthcare. For example, they could actively repress thoughts about symptoms, using various psychological avoidance-based coping mechanisms. It is plausible for some of these mechanisms to be cultivated more strongly in cultures that emphasize a greater gender differences between the sexes, as has been noted for Russia.(70, 71) To counterbalance this, the spouses of married men may pick up on changes to behaviour, to infer changed latent symptomology and biology, prompting inquisitive questions and ultimately urging their partners to seek healthcare. Men who live alone, but who have frequent contact with friends or relatives, may face a similar reaction, particularly to open-ended questions like *"How are you?"*. These could all serve as prompts which effectively disrupt the avoidance-style coping mechanism.

Secondly spouses, friends and relatives may all impart practical health-related knowledge (i.e. *"I know a man who had the same symptoms and this is what he did…"*) or knowledge about accessing healthcare (*"If you say the following to your GP, they will refer you more quickly to a specialist"*).

Third, in countries where healthcare requires out-of-pocket expenditure (e.g. as is the case for some countries in Eastern Europe), then friends and family may be able to provide the necessary funds to receive care. This may be particularly important for people on a lower income, so an *income*social support interaction* effect may be anticipated.

Fourth, it is plausible that men who eat dinner alone are more likely to cook quickly, relying on more processed and less healthy foods. In contrast, those who cooking communally with friends and family may be more likely to prepare a more time-consuming meal, which may be healthier. Bearing in mind that most events in my study occurred among participants in their 70s, it is also possible for men to have focused on

working roles throughout their life, and they may simply lack the basic cooking skills, to cook healthy food by themselves.

However, none of these four mechanisms above could account for why even among women, living alone, having limited contact with friends, having limited contact with relatives, as well as depression all seemed to exert independent effects. As I discussed in the introduction (section 1.1.5.1), it is plausible for the tendency and desire to form social relationships to be a more fundamental and innate property of the human condition, whose purpose may stretch beyond merely providing and receiving functional aid for those in need. If indeed such needs are so-called “hard wired” into the human brain, then people who are socially isolated may experience a range of emotional, cognitive and behavioural effects that are currently little understood. These in turn may have downstream consequences on a range of bodily systems (for example on altered inflammation, metabolic rate and/or the balance between anabolism-catabolism) which have not yet been uncovered.

4.2.2.4. International differences

The association between education and CHD varies somewhat between countries, in a pattern that typically follows the country's stage of socioeconomic development. It is plausible that, in European countries from 1700 to 1900, people with higher education (or other measure of higher SES) had a higher risk of developing CHD when compared to people with less education. Some of this may reflect survivor bias in settings where average life expectancy was between 30 to 40 years. People with less education were more likely to die from infectious disease, accident or injury before atherosclerotic processes had time to manifest. Accordingly, more educated people who had survived these earlier risks were more likely to live long enough to develop CHD.

Nonetheless, even after adjustment for age those with longer education may still have had a higher risk of developing CHD. This may have been since those with greater education were more likely to have been able to afford lifestyles that we now know are adverse cardiovascular risk factors: sedentary behaviour, a diet high in saturated fat and low in plant intake, smoking and alcohol excess. Some lower income countries today still show the same pattern. High SES people in non-Western settings today may still aspire to adopt Western lifestyles and unhealthy behaviours, particularly by shifting away from traditional food patterns towards more modern but less healthy diets, marked by greater BMI.

As countries further continue their socioeconomic development, it is typical for high SES people to see such unhealthy behaviours as maladaptive. Accordingly, middle-income countries tend to show flat educational gradients in CVD.

Following further socioeconomic development, the previously seen gradient tends to reverse in the opposite direction, whereby those with lowest education are now those with highest risk of CVD. These gradients may initially become particularly steep (as is seen in Eastern Europe, and especially Russia). Following this and with sustained socioeconomic

development towards high-income countries, socioeconomic gradients tend to persist but flatten somewhat (e.g. Russia → Baltic States → Central Europe → Scandinavia and Southern Europe (12, 13)).

Similar to the description in the paragraph above, the association between education and CHD in Western Europe may have changed somewhat over time. A range of forces may be in operation, causing sometimes opposite effects. For example, Western economies have become increasingly oriented around the service sector, where educational attainment may plausibly be a more important factor in determining economic outcomes and social position. This may widen educational gradients in health. Of particular interest is to think about what happens when the proportion of people who are employed in manual occupations falls over time. Under the hypothesis that it is the manual work itself that causes CHD, and also assuming that the hazard of this should gradually fall over time on account of advances to occupational health & safety, then one could expect the magnitude of association between occupation and health to fall, over time. However, under the hypothesis that manual work does not cause CHD, but is merely a marker of poor systemic fitness, then one could expect the magnitude of association between occupation and health to increase over time, as the pool of people in manual occupations become increasingly concentrated among those of lowest fitness. While there are few studies that have directly investigated this, there is a little support for the former hypothesis.(141) This is consistent with the theory that occupation may be over time partially displaced by other SES measures, such as education.

Assuming that education is causal for CHD (as discussed in earlier chapters), then the fact that educational opportunities and attainments have increased over time, may in itself have been a force which could have reduced health inequalities. However, evaluating this empirically would be difficult because of long lag effects of unclear duration, as well as how this shift over time is overlaid with other important period effects. Western Europe has seen notable political shits in the last century. The

period 1930-1980 was associated with greater public and political concern over socioeconomic inequalities being too steep, resulting in more redistributive policies (for example, inheritance taxes were raised to 80% in the UK and US, and stayed there throughout 1950-1980). It is plausible that this political shift could have also narrowed health inequalities. Such a health effect could have occurred immediately, or alternatively if considering 50 year lag times, by as late as 2030. After this initial shift towards the political left, the following period (1980-2018) has been marked by increasing socioeconomic inequality. It is plausible that this may increase health inequality, and under lagged phenomena these may continue to exert their effects until 2070. To summarize, Western European countries have seen considerable period effects over the last century. These include a shifting away from manual occupations, greater educational attainment, an initial period of reducing socioeconomic inequality, followed by a period of increasing socioeconomic inequality. While it is plausible for these changes to have altered the magnitude of health inequalities, it is difficult to make empirical attributions as such, on account of many of these changes occurring simultaneously and with different lagged effects.

Returning to the topic of explaining current international differences, then I am not aware of other cohort studies that have used a standardized protocol to investigate why cardiovascular mortality is so dramatically high in Russia when directly compared to elsewhere. Ecological data from the MONICA study suggested that 0-50% of the variations across time and place over Europe and Russia might be attributable to differences in conventional risk factors, supported by studies of smoking and alcohol in Russia.(167, 170) Using different methods 30 years later, I find that conventional as well as psychosocial factors do not account for much of the difference between Central European and Russian cohorts. Colleagues from the HAPIEE study have previously shown that two other hypothesised factors, alcohol(172) and dietary factors,(161) made only minor contribution to explaining the inter-cohort differences in mortality.

After considering the results of the current study and the wider literature, I am unable to clarify the reasons for the very high cardiovascular mortality in Russian men. On the one hand, Russia appears to have always had historically higher mortality rates, lending thought to the hypothesis of there being either latent genetic vulnerability, or cultural indifference for socio-political organizations that drive comparatively low performance for health outcomes. It is also possible that international differences in access to healthcare, in healthcare seeking behaviour, or in the quality of care received, may help to clarify this question. Somewhat problematically, any such factors would need to have a large differential impact on male mortality, but only a small differential impact on female mortality. Another area of enquiry might be gender norms and expectations. For example, women in Russia retire 5 years earlier than men. It is plausible that gender roles are more extremely polarized in such countries, which ultimately harms the health of men more than women.(71)

4.2.3. Implications for clinicians and policymakers

Similarly to elsewhere, psychosocial and socioeconomic factors are powerful predictors of cardiovascular mortality in Eastern Europe, and many of these mechanisms are probably causal. If conventional cardiovascular risk factors were to dramatically improve, then this would no doubt have large benefits on preventing cardiovascular disease. This may also attenuate socioeconomic gradients in cardiovascular disease to some degree. However, it is plausible that the vast majority of this may persist. The available body of evidence suggests that cardiovascular health could further improved, following the implementation of programs and policies that seek to:

- increase educational attainment (e.g. by raising the school leaving age to 18, or by fostering uptake of tertiary education);
- improve material conditions and lower unemployment (e.g. greater economic productivity and/or income redistribution);
- lower depression (e.g. introducing well-being literacy programmes to children and/or adults);
- provide more time for social relationships (e.g. awareness raising about the health benefits for example by integrating these into routine medical assessments; reducing total time spent at work; improving relationship literacy and skills)

4.2.4. Implications for future research

First, I have examined in earlier chapters the causality of the association from education to CHD. Similar causal investigations are now warranted for the other three independent sets of risk factors identified above. Material conditions are difficult to instrument genetically, but may be possible to study using the twin study or econometric designs. GWAS studies are emerging for depression, allowing future researchers to use Mendelian randomization to evaluate if depression causes CHD. Along similar lines, emerging GWAS studies on social isolation will allow Mendelian randomization analyses to be conducted on CHD and other health behaviour outcomes. It would be particularly useful to perform gender-specific analyses (requiring the exposure GWAS to be gender specific, and the outcome GWAS to be gender specific), in order to better understand any gender-specific effects.

Second, assuming that these pathways are causal, then the question remains as to *"What mechanisms account for the residual association from education, material conditions, depression and social isolation to CVD?"* Identifying these may yield novel interventional approaches, which can be used in addition to the upstream interventions that I suggested in section 4.2.3. To better examine behaviours around seeking for healthcare, analyses could be constructed on electronic healthcare record data linked to marriage records. These linked data could be evaluated, to see if those men who are married present for healthcare with milder disease, when compared to those men who are not married. If found, tailored programmes could be constructed to try to mitigate this effect, for example by prioritizing unmarried people in cardiovascular screening programmes such as the *NHS Health Check*. Where dietary behaviours are the putative mediator, epidemiologists experienced with psychosocial analyses could collaborate closely with nutritional epidemiologists in explaining CVD outcomes. Although much existing observational cohort data can support such analysis, limited research to date has directly addressed this level of complexity. For metabolomic mediators, one could

screen between 100 to 225 plasma metabolomic outcomes (as have been measured by nuclear magnetic resonance techniques for a number of cohorts) against a shortlist of socioeconomic or psychosocial exposures. In the case of 225 markers, the conventional $P<0.05$ threshold could be Bonferroni-corrected to $P<0.0002$. Such an analysis may require collaboration with multiple cohorts, to be both sufficiently powered and investigate external replication.

Third, better understanding of international differences may be furthered by a formal age-period-cohort analysis. These could be performed on each of the countries of Europe, and then cross-examined to search for any commonalities (or divergences) across the East-West axis. Another approach would be an econometric analysis. Approximately 20 Western and Easter European countries could be selected, to show up to 40 years of data each (whereby one data point is one country's average, for that year). The outcome variable could be CVD mortality, and researchers could compare simple and multivariable models to identify those risk factors that are strongest in attenuating the coefficients for country-dummies (and time-dummies). The top risk factors identified may or may not be causal, and would then need careful interpretation. A large obstacle to this agenda is the current absence of data availability to the granularity of annual population-level data, for many countries. As such, it may take many years if not decades until this analysis is feasible. Researchers interested in this may wish to collaborate with the *Global Burden of Disease* project(178, 200), as this project is also attempting to collate similar data.

4.2.5. Conclusion

To conclude, psychosocial and socioeconomic factors are powerful predictors of cardiovascular mortality in Eastern Europe, similarly to elsewhere. For some risk factors, such as unemployment, these associations appear stronger than elsewhere, indicating potential areas for policy intervention. Surprisingly, most of the associations between psychosocial risk factors and CVD were attenuated by only a small amount when adjusted for each other. This suggests that there may be a lot of nuanced pathways which could be relatively independent of one another, although I am aware that the causality of these associations requires further confirmation. The massive burden of cardiovascular mortality seen in Russia remains largely inexplicable.

4.3. Prediction

4.3.1. Summary of key findings

My study of risk prediction aimed to first evaluate performance of the original SCORE risk prediction model. Second, to derive a conventional CVD risk prediction model for use in Eastern European countries, using contemporary data. Third, to see if the addition of predominantly psychosocial and socioeconomic variables can improve model performance.

I found the existing SCORE model to be suboptimal in four contemporary Eastern European cohorts. In the derivation data it overpredicted risk, resulting in poor *cost-effectiveness*, while in the validation data it under predicted risk, resulting in poor *clinical effectiveness*. Recalibrating the original SCORE model and improving its mathematical format led to substantial improvements to *calibration, discrimination, binary NRI* and *clinical cost-effectiveness* in the derivation and validation datasets. The subsequent addition of seven novel cardiovascular risk factors led to further improvements to *discrimination, clinical effectiveness*, *binary NRI* and *Net Benefit,* in derivation and validation datasets. These findings were robust to sensitivity analyses that varied the clinical interventional threshold from 2% to 6%.

Taken together, the final model outperformed the original SCORE model in terms of *discrimination* (change in C-statistic of 0.06 and 0.03 in both the derivation and validation datasets, respectively), as well as *binary NRI* (0.07 and 0.14, respectively). *Cost-effectiveness*, the main problem in the derivation dataset, improved there by 39%. *Clinical effectiveness*, the main problem in the validation dataset, improved there by 38%. Altogether, this suggests that the final models appear to be superior to the original SCORE model, for each of the four countries studied.

4.3.2. Strengths and limitations

The main strength of this study was the use of an independent cohort from a different country for externally validating the two newly derived models. To the best of my knowledge, no previous study has added more than one socioeconomic or psychosocial factor to create a new prediction model, and validated its performance using extern data. Furthermore, as the derivation and validation datasets differed substantially in terms of their recruitment methods, urban/rural coverage, and data collection methods, this increases the likelihood that these findings may also replicate elsewhere.

Second, the derivation dataset had an appropriate number of events per variable, thereby reducing the risk of overfitting in the newly derived models. Third, all seven newly added risk factors are non-invasive and cheap to collect via self-report, thereby increasing their potential for widespread use across the world. Fourth, continuous variables were modelled appropriately, including the inclusion of quadratic functions for BMI and cholesterol. I am not aware of previous studies that have used a quadratic risk function for cholesterol.

Fifth, I used a comprehensive array of state of the art methods to thoroughly evaluate model performance. In particular, few cardiovascular risk prediction papers have attempted to combine the relative harms and benefits of intervention into an appropriately weighted metric, such as *Net Benefit* or *cost-effectiveness*. These have the potential to make risk prediction research more accessible, understandable and useful to a range of wider readers across clinical and public health specialties.

This study had some limitations. First, although the external validation dataset was sufficiently large to demonstrate that change in some outcomes (like the *Binary NRI*) was unlikely to happen by chance alone when comparing the original SCORE against the final HAPIEE model, power nonetheless appears to have been too small to assess the role of

chance in the intermediate stage (the recalibrated model 2, which was not augmented with new risk factors). To counteract this, it would have been helpful to additionally assess validation in the validation dataset that included imputed validation data. However, statistical packages for these analyses were readily available in neither STATA nor R, precluding me from performing these additional analyses.

Second, most cohort studies are subject to response rate bias, which also extends to my derivation and validation cohorts. This can be quite substantial, for example participants from the UK BioBank study have about half as many events when compared to those from the general UK population,(278) and it is likely that a similar healthy volunteer effect also appeared in my derivation data. This means that when I derived new models from the study participants, then I will make unrealistic expectations that the underlying population will be quite healthy (while in reality they are much more likely to develop CVD).

Nonetheless, this problem appears to be curiously offset by a similar reduction in the true event rate owing to the secular period effect. For example, the national mortality data on CVD mortality in Poland and the Czech Republic was around 350 per 100'000 in 2007, the year when most events occurred in my derivation data (See in the Methods section *"Figure 27. WHO mortality rates for cardiovascular disease in Poland and the Czech Republic."*). It may take around 10 years for national mortality offices and the cohort study's research team to collect and process such event data, derive a new clinical prediction model, publish this through peer-review and disseminate the model to policymakers and practitioners. It would then take another 10 years from implementation until the population being intervened upon start to develop event rates themselves (as most CVD models are operationalized with a 10-year follow-up time). If the current secular trends in WHO data are extrapolated until 2027, then this suggest that event rates may be around half of those seen in 2007. So if I apply my "unrealistically healthy model" derived in 2007 to the underlying population in 2007, then it will underestimate by around 50%.

However, if I apply my model to the underlying population in 2027, then it may estimate accurately. Hence by a coincidence, the healthy volunteer effect appears to be largely offset by unavoidable delays and corresponding period effects.

It may however be unrealistic to imagine that both of these two biases operate at similar magnitudes in each of the four countries that I studied. Hence my third limitation begins with the observation that the Central and Eastern European region has within it a very diverse and complex pattern of cardiovascular disease mortality. Much like the original SCORE paper, this current study has overly simplified this complexity by dichotomizing the region into two halves (low- and high-risk). This is a relatively crude approximation, which is likely to lead to suboptimal performance for countries in the middle, such as the Baltic States. It is possible that my (high-risk) model would over predict risk in the general population of Estonia for the year 2028, and this miscalibration has been temporarily masked by the fact that the healthy volunteer effect is unusually small in Estonian BioBank study (on account of this cohort having a more complex sampling method that more proactively recruited participants from healthcare services were would not be as healthy).

Ideally, such country-specific issues of miscalibration could be partly addressed by adapting the baseline hazard from each “off the shelf” model to the time and country context where it is being applied (such as “2027 Estonia”), by quantitatively triangulating and calibrating the baseline hazard from WHO mortality data. Such an approach was first tried by the GLOBORISK project, which was published in parallel to the writing of this thesis. However, as noted in my introduction chapter 1.4.3.2, the GLOBORISK model too may suffer from limitations of its own. Moreover, even if my models do have limitations with calibration, these do not distract from the overall finding from my analyses: that augmenting an existing model with 7 new CVD risk predictors led important improvements in model performance as measured by *discrimination, clinical effectiveness*, *binary NRI* and *Net Benefit.*

Fourth, the various performance metrics presented are difficult to reduce to a few key measures (despite the way that they occasionally appear to covary with one another), and furthermore these metrics present some internal inconsistencies. For example, the *Net Benefit* metric improved in the validation data, when comparing the original SCORE against the final HAPIEE-SCORE, but such improvement was not seen in the derivation data. This appears to have happened since the *Net Benefit* measure heavily rewards over prediction, which was rife when the original SCORE was used on the derivation data. Similarly, the results provided by the *Net Benefit calculations* (where I manually divided the NRI seen in controls by 20, tables 16-21) do not always match the *formal Net Benefit* calculated by the automated Decision Curve Analysis package. Furthermore, these two do not always match the calculation I performed to obtain *Net Clinical Effectiveness*, where I instigated my own QALY-weights from the wider literature. Although the latter of these three makes a couple of different assumptions, the theoretical similarity of all three methods made me expect a greater degree of convergence between their findings. As this theoretical approach to measuring performance is very new, I was unable to clarify to what degree these discrepancies may have arisen due to chance (on account of small marginal differences) or other reasons.

Fifth, it is possible that the process of augmenting EastEur-SCORE into the final HAPIEE-SCORE may have created a degree of overfitting. This can be seen in the drop in ΔC between the derivation (0.06) and validation (0.03) cohorts, as well as in the tendency for the final score to over predict risk in the validation set may have caused a drop in clinical cost-effectiveness in the validation data. Future work could seek to minimize this by performing shrinkage of the beta-coefficients, before the model is finalized. Additional benefits may be obtained by deriving the models on a dataset where missing data have been multiply imputed. However, such software is not readily available and missingness in the derivation data was low.

4.3.3. Comparison with previous studies

4.3.2.1. Recalibration

The newly recalibrated model I present (EastEur SCORE) appears to be superior to the original SCORE model. This is consistent with previous reports that found the SCORE model to perform poorly when validated with data from the HAPIEE study.(279) In some instances, calibration errors were so large as to show only one event, instead of the six that were predicted by SCORE.

The European Society of Cardiology, who endorsed the original SCORE prediction model, have recommended that SCORE be recalibrated for each particular country in question, in collaboration with national societies of cardiology. Of the 25 countries allocated to the “low risk” region (predominantly from Western Europe) the heartscore.org website lists 7 countries (i.e. 28%) who have recalibrated SCORE for their country context. This does not include countries such as the UK, who have derived their own cardiovascular risk prediction model without reference to SCORE. Of the 25 countries allocated to the “high risk” region (predominantly from Eastern Europe), the heartscore.org website lists 9 countries (i.e. 36%) who have recalibrated SCORE for their country context. For example, the Polish Cardiac Society has recalibrated SCORE twice (first in 2007, and later updated in 2015) [Personal communication with Andrzej Pajak]. However, none of these recalibrated models have been published in English. This absence is present both on peer-reviewed journals, as well as elsewhere (such as the websites of heartscore.org or the European Society of Cardiology). The recalibrated models are freely available online to make individual risk predictions for one participant. However, it is impractical to use these if a researcher like myself wishes to evaluate prediction performance for a large cohort of participants. For such a purpose, the original SCORE models remain the default option. Another promising default model is the GLOBORISK model.(199)

However, this model combined fatal and non-fatal CVD end points, thereby increasing international heterogeneity in performance.

I am aware of only one prior example, in the English-speaking peer-reviewed literature, where the original SCORE has been compared against a new model. Jdanov et al. used cohort data from Moscow and St. Petersburg, collected from 1975-2001, to derive a new score for Russia (SCORE-MoSP).(203) However, this study did not adjust for the chaotic fluctuations in baseline rate during this period. Importantly, the model they created has not been externally validated, leaving open the possibility that they may not work for other cohort studies or real-life clinical encounters, with a different case mix. Future studies could compare the performance of the original SCORE against the SCORE-MoSP using validation data from Novosibirsk, such as the MONICA or the HAPIEE study.

To summarize, the newly derived “EastEur SCORE” appears to be the first example in the English-speaking peer-reviewed literature, where a new cardiovascular risk prediction model has been derived for an Eastern European country, this has been validated with external data, and shown to be superior against the Original SCORE model.

4.3.2.2. Augmentation

Previous studies that tried to improve cardiovascular risk prediction by adding any kind of new predictors have typically found this to be a difficult endeavour. Derivation studies have sometimes reported improvements to discrimination as high as ΔC=0.08. However, these have rarely been matched in external validation studies, where ΔC greater than 0.01 to 0.02 has rarely been achieved. In my study, ΔC of 0.022 in derivation and 0.014 in validation is broadly consistent with this range. As one typical comparison, Abraham et al. added 49 310 genome-wide SNPs (strongly and weakly associated with CHD), and reported ΔCs of 0.011 to 0.017 (depending on the particular cohort and analyses performed). Few of these studies have reported benefits in terms of clinically meaningful reclassification, while the clinically less relevant Continuous NRI is often

quoted. In the aforementioned genomic augmentation, clinical NRI was smaller (0.25 to 0.37) than what I report in my study (0.44 to 0.53). This suggests that the addition of simple behavioural and psychosocial questions may be at least as useful as the best data across the entire genome, if not better.

Indeed, the literature on augmenting CVD risk prediction models is predominantly focused on technology-intensive measures, such as blood-based ‘omics predictors as well as various imaging modalities. In comparison, the literature on augmentation following the addition of behavioural and/or psychosocial questions is very sparse. About 10 years ago, when well-known scores like ASSIGN and QRISK were developed (for Scotland and England, respectively) then these made a number of changes in comparison to their SCORE or Framingham benchmarks that they tried to improve. These included area-level psychosocial status and BMI. This means it is hard to disentangle what part of their improvement came from recalibration, and what part from the addition of novel risk factors. This is something that my study was able to separate into two distinct parts.

There are at least 10 other studies that have investigated the benefit of adding psychosocial/socioeconomic risk factors (detailed in Table 2 in the introduction chapter). Some of the early studies were underpowered, where a low event-to-parameter ratios probably led to considerable overfit, the possibility of chance findings, and lower probability of succeeding theoretically, were these ever taken to external validation.(18, 212, 217). As such, although these studies reported very large improvements to discrimination (ΔCs of 0.02 to 0.04), it is difficult to tell if this reflects a true improvement or a chance finding from overfit.

More recent studies with better statistical power have reported smaller benefits to discrimination (ΔCs of 0.004 to 0.01), after including one to three new predictors.(12, 110, 214, 218-221). My results are consistent with these, since I demonstrate a much larger improvement as a result of

adding seven new predictors. As some of these predictors are partly collinear with each other, this explains why the ΔC in my study was not as large as 7 * (0.004 to 0.01). Instead, the magnitude is in the middle of the range expected, based on the previous studies on some of these predictors in isolation. This is also broadly consistent with the magnitude of improvements seen in *Continuous NRI* (0.03 to 0.18 in previous studies; 0.44 to 0.53 in my study) and *Categorical NRI* (0.03 to 0.18 in previous studies, 0.03 to 0.06 in my study). To the best of my knowledge, no previous study has added psychosocial or socioeconomic risk factors, and evaluated changes using a *Net Benefit, Number Needed to Screen,* or *clinical cost-effectiveness*.

Perhaps most importantly, no previous study has to the best of my knowledge demonstrated an added benefit of adding a range of psychosocial, socioeconomic and behavioural factors, which has then been externally validated. Perhaps my study could also inspire others to consider the potential benefit of adding such risk factors. If similar improvements could be generated in other settings, then this might create potential for such factors to become a part of CVD risk prediction models across the world. This may be particularly useful for low-resource contexts (e.g. India and possibly also China and Russia) which together have the largest burden of cardiovascular disease (figure 51).

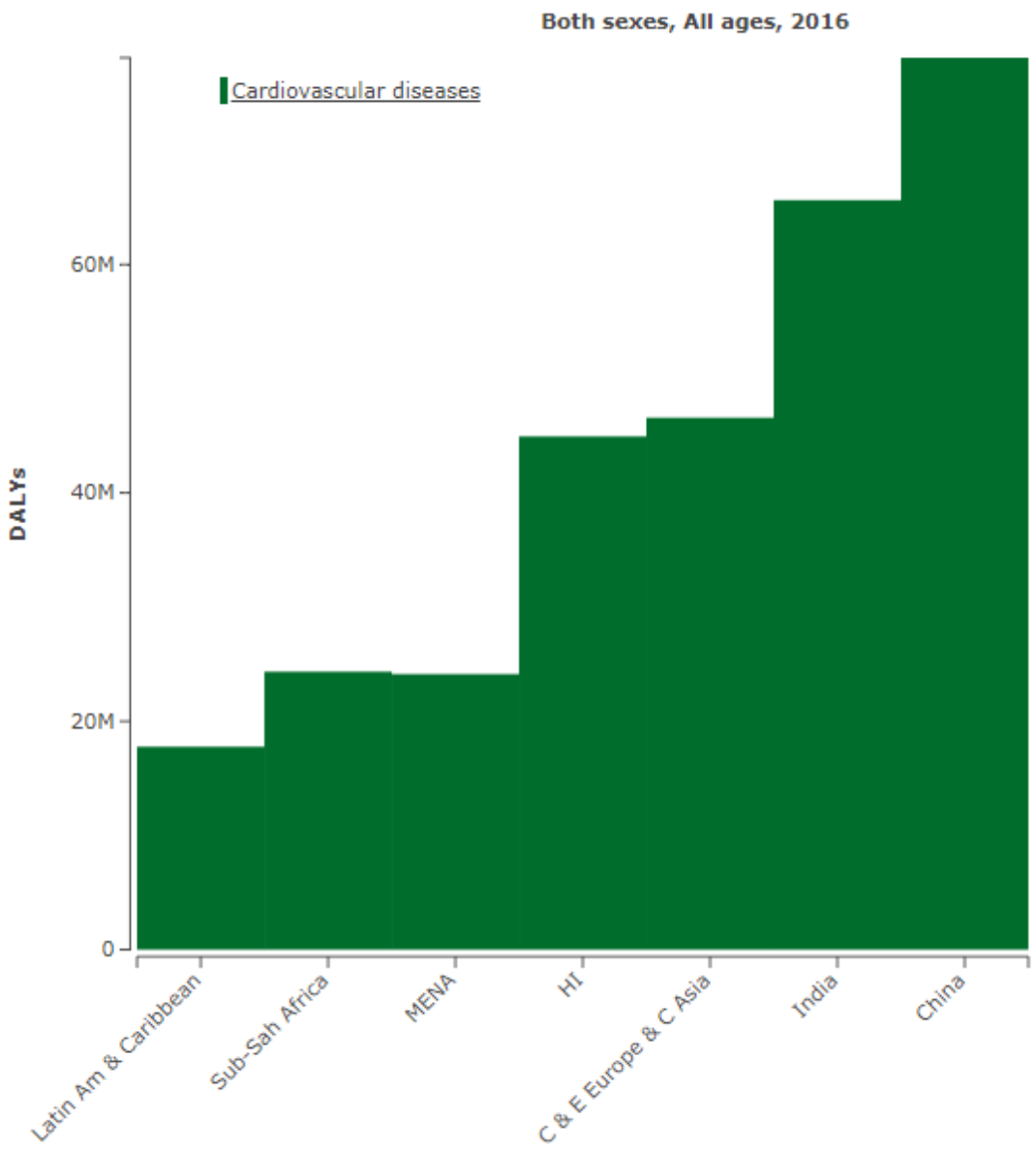


Figure 51. Total burden of cardiovascular disease in selected world regions and countries*. DALY = Disability Adjusted Life Year; MENA = Middle East and North Africa; HI = High Income countries; C & E = Central and Eastern. Data are from (200), which originally listed 7 regional areas. For the graph above, two countries from the two largest areas (India and China) were presented, thereby excluding data from the other Asian countries. Data from other world regions are all included in this illustration.*

4.3.4. Implications for clinicians, policymakers and for future research

Altogether, these findings of the utility of non-invasive predictors, supports a wider reconsideration of how and where personalized prediction services are delivered. To date, this has required all participants to attend a face-to-face appointment at a medical facility. Instead, a multi-stage approach could be considered in the future, whereby initial pre-screening is done with existing records, early screening is done virtually with online self-report data, and final-stage screening that involves face-to-face and more invasive data collection (e.g. cholesterol and blood pressure) is recommended only for those of intermediate or high risk.(280, 281). This could lower programme cost and create opportunities to target interventions and reduce health inequalities in an increasingly personalized way. In particular, as existing cardiovascular screening programmes have not yet demonstrated their effectiveness, such reconfigurations to service delivery may open up new ways of thinking about what an effective and cost-effective CVD screening programme could look like.

5. Conclusions

People whose genetic predisposition causes them to attain more education also have a large reduction in their subsequent risk of CHD. In conjunction with the wider evidence, this suggests that policy changes which increase education are more likely than not to prevent a large fraction of CHD. This may be partly explained by changed smoking and to a lesser degree some of the other conventional CVD risk factors. As such, interventions that reduce smoking are likely to be among the most effective in terms of lowering health inequalities. However, conventional CVD risk factors are unlikely to account entirely for the putative causal mechanism from education to CHD, which is one potential area of future research.

The large international difference seen between Eastern and Western European countries remains largely inexplicable. Fresh hypotheses are needed to investigate the causes of these. My own personal observations have made me wonder whether differences in the quality of healthcare may play a role in this.

Most of the associations between a single socioeconomic or psychosocial risk factor and CVD mortality attenuate by only a small amount, after adjustment for other socioeconomic and psychosocial factors, or when adjustment for conventional CVD risk factors. This suggests that a range of nuanced pathways may associate various socioeconomic and psychosocial factors to CVD, and that these could be relatively independent of one another. If causal, this suggests that a range of changes to depression, social support, material conditions and unemployment are all likely to cause considerable CVD benefit. Furthermore, then conventional CVD risk factors are unlikely to account for the majority of these mechanisms. If future research could identify their underlying mechanisms, then this may suggest novel avenues for intervention that could assist in lowering health inequalities.

Adding predominantly socioeconomic/psychosocial risk factors to risk prediction models led to substantial and clinically meaningful increases in prediction performance. The new model (called HAPIEE-SCORE) could now be evaluated in clinical primary care settings (if necessary, with further recalibration beforehand). For example, some Eastern European countries might be well suited to conduct a cluster randomized trial of a CVD screening programme, since the control arm here would receive comparatively little baseline intervention. As the region has high event rates, such a study would be well-powered and cost-effective.

Overall this thesis underscores the notion that although socioeconomic and psychosocial risk factors may not be that useful for understanding international differences, they do appear to be highly relevant to clinicians, public health professionals, and to those interested in the co-benefits from education to health. I hope that these findings will assist in translating such knowledge, which to date has been developed primarily between epidemiologists, more widely into the everyday practice of a wider body of health professionals. I hope that this offers some assistance towards the goal of eradicating heart disease, as well as the large social inequalities with which it currently manifests.

References

1. Davey Smith G, Hart C, Hole D, MacKinnon P, Gillis C, Watt G, et al. Education and occupational social class: which is the more important indicator of mortality risk? J Epidemiol Community Health. 1998;52(3):153-60.
2. Vandenheede H, Vikhireva O, Pikhart H, Kubinova R, Malyutina S, Pajak A, et al. Socioeconomic inequalities in all-cause mortality in the Czech Republic, Russia, Poland and Lithuania in the 2000s: findings from the HAPIEE Study. J Epidemiol Community Health. 2014;68(4):297-303.
3. Allen L, Williams J, Townsend N, Mikkelsen B, Roberts N, Foster C, et al. Socioeconomic status and non-communicable disease behavioural risk factors in low-income and lower-middle-income countries: a systematic review. The Lancet Global Health. 2017;5(3):e277-e89.
4. Leinsalu M, Vågerö D, Kunst AE. Estonia 1989–2000: enormous increase in mortality differences by education. International journal of epidemiology. 2003;32(6):1081-7.
5. Hinkle LE, Jr., Whitney LH, Lehman EW, Dunn J, Benjamin B, King R, et al. Occupation, education, and coronary heart disease. Risk is influenced more by education and background than by occupational experiences, in the Bell System. Science. 1968;161(3838):238-46.
6. Shekelle RB, Ostfeld AM, Paul O. Social status and incidence of coronary heart disease. Journal of chronic diseases. 1969;22(6):381-94.
7. Kitagawa EM, Hauser PM. Differential mortality in the United States: A study in socioeconomic epidemiology. 1st ed. Cambridge, Massachusetts: Harvard University Press; 1973.
8. Manrique-Garcia E, Sidorchuk A, Hallqvist J, Moradi T. Socioeconomic position and incidence of acute myocardial infarction: a meta-analysis. J Epidemiol Community Health. 2011;65(4):301-9.
9. Vathesatogkit P, Batty GD, Woodward M. Socioeconomic disadvantage and disease-specific mortality in Asia: systematic review with meta-analysis of population-based cohort studies. J Epidemiol Community Health. 2014;68(4):375-83.
10. Khaing W, Vallibhakara SA, Attia J, McEvoy M, Thakkinstian A. Effects of education and income on cardiovascular outcomes: A systematic review and meta-analysis. European Journal of Preventive Cardiology. 2017:2047487317705916.
11. Veronesi G, Ferrario MM, Kuulasmaa K, Bobak M, Chambless LE, Salomaa V, et al. Educational class inequalities in the incidence of coronary heart disease in Europe. Heart. 2016;102(12):958-65.
12. Ferrario MM, Veronesi G, Chambless LE, Tunstall-Pedoe H, Kuulasmaa K, Salomaa V, et al. The contribution of educational class in improving accuracy of cardiovascular risk prediction across European regions: The MORGAM Project Cohort Component. Heart. 2014;100(15):1179-87.
13. Mackenbach JP, Stirbu I, Roskam A-JR, Schaap MM, Menvielle G, Leinsalu M, et al. Socioeconomic inequalities in health in 22 European countries. New England Journal of Medicine. 2008;358(23):2468-81.
14. Bobak M, Murphy M, Pikhart H, Martikainen P, Rose R, Marmot M. Mortality patterns in the Russian Federation: indirect technique using

widowhood data. Bulletin of the World Health Organization. 2002;80(11):876-81.
15. Bobak M, Murphy M, Rose R, Marmot M. Determinants of adult mortality in Russia: estimates from sibling data. Epidemiology. 2003;14(5):603-11.
16. Malyutina S, Bobak M, Simonova G, Gafarov V, Nikitin Y, Marmot M. Education, marital status, and total and cardiovascular mortality in Novosibirsk, Russia: a prospective cohort study. Annals of epidemiology. 2004;14(4):244-9.
17. Perlman F, Bobak M. Socioeconomic and behavioral determinants of mortality in posttransition Russia: a prospective population study. Annals of epidemiology. 2008;18(2):92-100.
18. Vikhireva O, Broda G, Kubinova R, Malyutina S, Pająk A, Tamosiunas A, et al. Does inclusion of education and marital status improve SCORE performance in Central and Eastern Europe and former Soviet Union? Findings from MONICA and HAPIEE Cohorts. PloS one. 2014;9(4):e94344.
19. Hafeman DM. Confounding of indirect effects: a sensitivity analysis exploring the range of bias due to a cause common to both the mediator and the outcome. American journal of epidemiology. 2011;174(6):710-7.
20. Blakely T, McKenzie S, Carter K. Misclassification of the mediator matters when estimating indirect effects. J Epidemiol Community Health. 2013;67(5):458-66.
21. OECD. Society at a Glance: OECD Social Indicators 2006 Edition. Paris: Organisation for Economic Co-operation and Development (OECD); 2007. https://stats.oecd.org/glossary/detail.asp?ID=7326.
22. Macleod J, Smith GD, Metcalfe C, Hart C. Is subjective social status a more important determinant of health than objective social status? Evidence from a prospective observational study of Scottish men. Social Science & Medicine. 2005;61(9):1916-29.
23. Nelson CP, Hamby SE, Saleheen D, Hopewell JC, Zeng L, Assimes TL, et al. Genetically determined height and coronary artery disease. N Engl J Med. 2015;372(17):1608-18.
24. Yngwe MÅ, Kondo N, Hägg S, Kawachi I. Relative deprivation and mortality–a longitudinal study in a Swedish population of 4, 7 million, 1990–2006. BMC public health. 2012;12(1):664.
25. Kondo N, Sembajwe G, Kawachi I, van Dam RM, Subramanian S, Yamagata Z. Income inequality, mortality, and self rated health: meta-analysis of multilevel studies. BMJ. 2009;339:b4471.
26. Pikhart H, Bobak M, Rose R, Marmot M. Household item ownership and self-rated health: material and psychosocial explanations. BMC public health. 2003;3(1):38.
27. Stelmach W, Kaczmarczyk-Chałas K, Bielecki W, Stelmach I, Drygas W. How income and education contribute to risk factors for cardiovascular disease in the elderly in a former Communist country. Public health. 2004;118(6):439-49.
28. Bobak M, Pikhart H, Rose R, Hertzman C, Marmot M. Socioeconomic factors, material inequalities, and perceived control in self-rated health: cross-sectional data from seven post-communist countries. Social science & medicine. 2000;51(9):1343-50.
29. Moser KA, Fox AJ, Jones D. Unemployment and mortality in the OPCS longitudinal study. The Lancet. 1984;324(8415):1324-9.

30. Roelfs DJ, Shor E, Davidson KW, Schwartz JE. Losing life and livelihood: a systematic review and meta-analysis of unemployment and all-cause mortality. Social Science & Medicine. 2011;72(6):840-54.
31. McLeod CB, Lavis JN, MacNab YC, Hertzman C. Unemployment and mortality: a comparative study of Germany and the United States. American journal of public health. 2012;102(8):1542-50.
32. Perlman F, Bobak M. Assessing the contribution of unstable employment to mortality in posttransition Russia: prospective individual-level analyses from the Russian longitudinal monitoring survey. American Journal of Public Health. 2009;99(10):1818-25.
33. Shkolnikov VM, Meslé F, Leon DA. Premature circulatory disease mortality in Russia: population-and individual-level evidence. Heart disease: environment, stress and gender. 2002:39-68.
34. Dupre ME, George LK, Liu G, Peterson ED. The cumulative effect of unemployment on risks for acute myocardial infarction. Archives of internal medicine. 2012;172(22):1731-7.
35. Jokela M, Pulkki-Råback L, Elovainio M, Kivimäki M. Personality traits as risk factors for stroke and coronary heart disease mortality: pooled analysis of three cohort studies. Journal of Behavioral Medicine. 2014;37(5):881-9.
36. Gallo WT, Teng H-M, Falba TA, Kasl SV, Krumholz HM, Bradley EH. The impact of late career job loss on myocardial infarction and stroke: a 10 year follow up using the health and retirement survey. Occupational and environmental medicine. 2006;63(10):683-7.
37. Meneton P, Kesse-Guyot E, Méjean C, Fezeu L, Galan P, Hercberg S, et al. Unemployment is associated with high cardiovascular event rate and increased all-cause mortality in middle-aged socially privileged individuals. International archives of occupational and environmental health. 2015;88(6):707-16.
38. Strully KW. Job loss and health in the US labor market. Demography. 2009;46(2):221-46.
39. Sullivan D, Von Wachter T. Job displacement and mortality: An analysis using administrative data. The Quarterly Journal of Economics. 2009;124(3):1265-306.
40. Bobak M, McKee M, Rose R, Marmot M. Alcohol consumption in a national sample of the Russian population. Addiction. 1999;94(6):857-66.
41. Pomerleau J, McKee M, Rose R, Haerpfer CW, Rotman D, Tumanov S. Hazardous alcohol drinking in the former Soviet Union: a cross-sectional study of eight countries. Alcohol & alcoholism. 2008;43(3):351-9.
42. Azarova A, Irdam D, Gugushvili A, Fazekas M, Scheiring G, Horvat P, et al. The effect of rapid privatisation on mortality in mono-industrial towns in post-Soviet Russia: a retrospective cohort study. The Lancet Public Health. 2017;2(5):e231-e8.
43. Riumallo-Herl C, Basu S, Stuckler D, Courtin E, Avendano M. Job loss, wealth and depression during the Great Recession in the USA and Europe. International journal of epidemiology. 2014;43(5):1508-17.
44. Durkheim E. Suicide: A study in sociology (JA Spaulding & G. Simpson, trans.). Glencoe, IL: Free Press(Original work published 1897). 1951.
45. Cobb S. Social support as a moderator of life stress. Psychosomatic medicine. 1976;38(5):300-14.

46. Cassel J. The contribution of the Social Environment to Host Resistance. American journal of epidemiology. 1976;104(2):107-23.
47. Berkman L, Krishna A. Social Network Epidemiolgy. In Social Epidemiology, 2nd Ed: Oxford University Press; 2014.
48. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. n engl j med. 2007;2007(357):370-9.
49. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. American journal of epidemiology. 1979;109(2):186-204.
50. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychological bulletin. 1985;98(2):310.
51. Karasek Jr RA. Job demands, job decision latitude, and mental strain: Implications for job redesign. Administrative science quarterly. 1979:285-308.
52. Johnson JV, Hall EM. Job strain, work place social support, and cardiovascular disease: a cross-sectional study of a random sample of the Swedish working population. American journal of public health. 1988;78(10):1336-42.
53. Bosma H, Marmot MG, Hemingway H, Nicholson AC, Brunner E, Stansfeld SA. Low job control and risk of coronary heart disease in Whitehall II (prospective cohort) study. Bmj. 1997;314(7080):558.
54. Kaufmann GM, Beehr TA. Interactions between job stressors and social support: Some counterintuitive results. Journal of Applied Psychology. 1986;71(3):522.
55. Vilhjalmsson R. Life stress, social support and clinical depression: a reanalysis of the literature. Social science & medicine. 1993;37(3):331-42.
56. Graeber D. Debt-updated and expanded: the first 5,000 years: Melville House; 2014.
57. McEwen BS. Stress, adaptation, and disease: Allostasis and allostatic load. Annals of the New York Academy of Sciences. 1998;840(1):33-44.
58. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS medicine. 2010;7(7):e1000316.
59. Ramsay S, Ebrahim S, Whincup P, Papacosta O, Morris R, Lennon L, et al. Social engagement and the risk of cardiovascular disease mortality: results of a prospective population-based study of older men. Annals of epidemiology. 2008;18(6):476-83.
60. Ikeda A, Iso H, Kawachi I, Yamagishi K, Inoue M, Tsugane S. Social support and stroke and coronary heart disease. Stroke. 2008;39(3):768-75.
61. Barth J, Schneider S, von Känel R. Lack of social support in the etiology and the prognosis of coronary heart disease: a systematic review and meta-analysis. Psychosomatic Medicine. 2010;72(3):229-38.
62. Kuper H, Marmot M, Hemingway H, editors. Systematic review of prospective cohort studies of psychosocial factors in the etiology and prognosis of coronary heart disease. Seminars in vascular medicine; 2002: Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4662.
63. Schmaltz HN, Southern D, Ghali WA, Jelinski SE, Parsons GA, King KM, et al. Living alone, patient sex and mortality after acute myocardial infarction. Journal of general internal medicine. 2007;22(5):572.
64. Stringhini S, Berkman L, Dugravot A, Ferrie JE, Marmot M, Kivimaki M, et al. Socioeconomic status, structural and functional measures of social

support, and mortality: The British Whitehall II Cohort Study, 1985–2009. American journal of epidemiology. 2012;175(12):1275-83.
65. Robles TF, Slatcher RB, Trombello JM, McGinn MM. Marital quality and health: A meta-analytic review. Psychological bulletin. 2014;140(1):140.
66. Kawachi I, Colditz GA, Ascherio A, Rimm EB, Giovannucci E, Stampfer MJ, et al. A prospective study of social networks in relation to total mortality and cardiovascular disease in men in the USA. Journal of Epidemiology & Community Health. 1996;50(3):245-51.
67. Eng PM, Rimm EB, Fitzmaurice G, Kawachi I. Social ties and change in social ties in relation to subsequent total and cause-specific mortality and coronary heart disease incidence in men. American journal of epidemiology. 2002;155(8):700-9.
68. Gallo LC, Troxel WM, Kuller LH, Sutton-Tyrrell K, Edmundowicz D, Matthews KA. Marital status, marital quality, and atherosclerotic burden in postmenopausal women. Psychosomatic medicine. 2003;65(6):952-62.
69. André-Petersson L, Hedblad B, Janzon L. Social support and behavior in a stressful situation in relation to myocardial infarction and mortality: who is at risk? Results from prospective cohort study “Men born in 1914,” Malmö, Sweden. International journal of behavioral medicine. 2006;13(4):340.
70. Hajdu P, Mckee M, Bojan F. Changes in premature mortality differentials by marital status in Hungary and in England and Wales. The European Journal of Public Health. 1995;5(4):259-64.
71. Watson P. Explaining rising mortality among men in Eastern Europe. Social Science & Medicine. 1995;41(7):923-34.
72. Rosenman RH, Brand RJ, Jenkins CD, Friedman M, Straus R, Wurm M. Coronary heart disease in the Western Collaborative Group Study: Final follow-up experience of 8 1/2 years. Jama. 1975;233(8):872-7.
73. Chida Y, Steptoe A. The association of anger and hostility with future coronary heart disease: a meta-analytic review of prospective evidence. Journal of the American college of cardiology. 2009;53(11):936-46.
74. Friedman M, Thoresen CE, Gill JJ, Ulmer D, Powell LH, Price VA, et al. Alteration of type A behavior and its effect on cardiac recurrences in post myocardial infarction patients: summary results of the recurrent coronary prevention project. American heart journal. 1986;112(4):653-65.
75. Eysenck HJ, Eysenck S. The Eysenck personality inventory. 1965.
76. McCrae RR, Costa PT. Validation of the five-factor model of personality across instruments and observers. Journal of personality and social psychology. 1987;52(1):81.
77. Digman JM. Five robust trait dimensions: Development, stability, and utility. Journal of personality. 1989;57(2):195-214.
78. Goldberg LR. An alternative" description of personality": the big-five factor structure. Journal of personality and social psychology. 1990;59(6):1216.
79. Jokela M, Batty GD, Nyberg ST, Virtanen M, Nabi H, Singh-Manoux A, et al. Personality and all-cause mortality: individual-participant meta-analysis of 3,947 deaths in 76,150 adults. American journal of epidemiology. 2013;178(5):667-75.
80. Nakaya N, Tsubono Y, Hosokawa T, Hozawa A, Kuriyama S, Fukudo S, et al. Personality and mortality from ischemic heart disease and stroke. Clinical and experimental hypertension. 2005;27(2-3):297-305.

81. Batty GD, Jokela M, Kivimaki M, Shipley M. Examining the Long-Term Association of Personality With Cause-Specific Mortality in London: Four Decades of Mortality Surveillance in the Original Whitehall Smoking Cessation Trial. American journal of epidemiology. 2016;184(6):436-41.
82. Luchetti M, Barkley JM, Stephan Y, Terracciano A, Sutin AR. Five-factor model personality traits and inflammatory markers: New data and a meta-analysis. Psychoneuroendocrinology. 2014;50:181-93.
83. Keltikangas-Järvinen L, Pulkki-Råback L, Puttonen S, Viikari J, Raitakari OT. Childhood hyperactivity as a predictor of carotid artery intima media thickness over a period of 21 years: the cardiovascular risk in young Finns study. Psychosomatic medicine. 2006;68(4):509-16.
84. Bandura A. The explanatory and predictive scope of self-efficacy theory. Journal of social and clinical psychology. 1986;4(3):359-73.
85. Krause N, Shaw BA. Role-specific feelings of control and mortality. Psychology and aging. 2000;15(4):617.
86. Dalgard OS, Håheim LL. Psychosocial risk factors and mortality: a prospective study with special focus on social support, social participation, and locus of control in Norway. Journal of Epidemiology & Community Health. 1998;52(8):476-81.
87. Stürmer T, Hasselbach P, Amelang M. Personality, lifestyle, and risk of cardiovascular disease and cancer: follow-up of population based cohort. BMJ. 2006;332(7554):1359.
88. van Oort FV, van Lenthe FJ, Mackenbach JP. Material, psychosocial, and behavioural factors in the explanation of educational inequalities in mortality in The Netherlands. Journal of Epidemiology & Community Health. 2005;59(3):214-20.
89. Surtees PG, Wainwright NW, Luben R, Khaw K-T, Day NE. Mastery, sense of coherence, and mortality: evidence of independent associations from the EPIC-Norfolk Prospective Cohort Study. Health Psychology. 2006;25(1):102.
90. Surtees PG, Wainwright NW, Luben R, Wareham NJ, Bingham SA, Khaw K-T. Mastery is associated with cardiovascular disease mortality in men and women at apparently low risk. Health Psychology. 2010;29(4):412.
91. Kozela M, Doryńska A, Stepaniak U, Szafraniec K, Topór-Mądry R, Misiowiec W, et al. Perceived control as a predictor of cardiovascular disease mortality in Poland. The HAPIEE study. Cardiology journal. 2015;22(4):404-12.
92. De Vogli R, Ferrie JE, Chandola T, Kivimäki M, Marmot MG. Unfairness and health: evidence from the Whitehall II Study. Journal of Epidemiology & Community Health. 2007;61(6):513-8.
93. Anda R, Williamson D, Jones D, Macera C, Eaker E, Glassman A, et al. Depressed affect, hopelessness, and the risk of ischemic heart disease in a cohort of US adults. Epidemiology. 1993:285-94.
94. Van der Kooy K, van Hout H, Marwijk H, Marten H, Stehouwer C, Beekman A. Depression and the risk for cardiovascular diseases: systematic review and meta analysis. International journal of geriatric psychiatry. 2007;22(7):613-26.
95. Rugulies R. Depression as a predictor for coronary heart disease: a review and meta-analysis. American journal of preventive medicine. 2002;23(1):51-61.

96. Bekke-Hansen S, Trockel M, Burg MM, Taylor CB. Depressive symptom dimensions and cardiac prognosis following myocardial infarction: results from the ENRICHD clinical trial. Psychological medicine. 2012;42(1):51-60.
97. Barefoot JC, Brummett BH, Helms MJ, Mark DB, Siegler IC, Williams RB. Depressive symptoms and survival of patients with coronary artery disease. Psychosomatic Medicine. 2000;62(6):790-5.
98. Kozela M, Bobak M, Besala A, Micek A, Kubinova R, Malyutina S, et al. The association of depressive symptoms with cardiovascular and all-cause mortality in Central and Eastern Europe: prospective results of the HAPIEE study. European journal of preventive cardiology. 2016;23(17):1839-47.
99. Hamer M, Batty GD, Seldenrijk A, Kivimaki M. Antidepressant medication use and future risk of cardiovascular disease: the Scottish Health Survey. European heart journal. 2010;32(4):437-42.
100. Labos C, Dasgupta K, Nedjar H, Turecki G, Rahme E. Risk of bleeding associated with combined use of selective serotonin reuptake inhibitors and antiplatelet therapy following acute myocardial infarction. Canadian Medical Association Journal. 2011;183(16):1835-43.
101. Glassman AH, O'connor CM, Califf RM, Swedberg K, Schwartz P, Bigger Jr JT, et al. Sertraline treatment of major depression in patients with acute MI or unstable angina. Jama. 2002;288(6):701-9.
102. Berkman LF, Blumenthal J, Burg M, Carney RM, Catellier D, Cowan MJ, et al. Effects of treating depression and low perceived social support on clinical events after myocardial infarction: the Enhancing Recovery in Coronary Heart Disease Patients (ENRICHD) Randomized Trial. JAMA: Journal of the American Medical Association. 2003.
103. Scherrer JF, Xian H, Bucholz KK, Eisen SA, Lyons MJ, Goldberg J, et al. A twin study of depression symptoms, hypertension, and heart disease in middle-aged men. Psychosomatic Medicine. 2003;65(4):548-57.
104. Rees K, Bennett P, West R, Davey Smith G, Ebrahim S. Psychological interventions for coronary heart disease. Cochrane Database Syst Rev. 2004(2):CD002902.
105. Linden W, Phillips MJ, Leclerc J. Psychological treatment of cardiac patients: a meta-analysis. European heart journal. 2007;28(24):2972-84.
106. Huppert FA, Whittington JE. Symptoms of psychological distress predict 7-year mortality. Psychological medicine. 1995;25(5):1073-86.
107. Russ TC, Stamatakis E, Hamer M, Starr JM, Kivimäki M, Batty GD. Association between psychological distress and mortality: individual participant pooled analysis of 10 prospective cohort studies. bmj. 2012;345:e4933.
108. Appels A, Höppener P, Mulder P. A questionnaire to assess premonitory symptoms of myocardial infarction. International journal of cardiology. 1987;17(1):15-24.
109. Kopp MS, Falger PR, Appels A, Szedmak S. Depressive symptomatology and vital exhaustion are differentially related to behavioral risk factors for coronary artery disease. Psychosomatic medicine. 1998;60(6):752-8.
110. Schnohr P, Marott JL, Kristensen TS, Gyntelberg F, Gronbaek M, Lange P, et al. Ranking of psychosocial and traditional risk factors by

importance for coronary heart disease: the Copenhagen City Heart Study. Eur Heart J. 2015;36(22):1385-93.
111. Boehm JK, Peterson C, Kivimaki M, Kubzansky LD. Heart health when life is satisfying: evidence from the Whitehall II cohort study. European Heart Journal. 2011;32(21):2672-7.
112. Roest AM, Martens EJ, de Jonge P, Denollet J. Anxiety and risk of incident coronary heart disease: a meta-analysis. Journal of the American College of Cardiology. 2010;56(1):38-46.
113. Phillips AC, Batty GD, Gale CR, Deary IJ, Osborn D, MacIntyre K, et al. Generalized anxiety disorder, major depressive disorder, and their comorbidity as predictors of all-cause and cardiovascular mortality: the Vietnam experience study. Psychosomatic Medicine. 2009;71(4):395-403.
114. Kumar P, Clark M. Clinical Medicine. 9th Ed. ed: Elsevier; 2017.
115. Keavney B, Danesh J, Parish S, Palmer A, Clark S, Youngman L, et al. Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'. International journal of epidemiology. 2006;35(4):935-43.
116. Wensley Fea. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. Bmj. 2011;342:d548.
117. Van Meurs JB, Pare G, Schwartz SM, Hazra A, Tanaka T, Vermeulen SH, et al. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. The American journal of clinical nutrition. 2013;98(3):668-76.
118. Yang Q, Köttgen A, Dehghan A, Smith AV, Glazer NL, Chen M-H, et al. Multiple genetic loci influence serum urate and their relationship with gout and cardiovascular disease risk factors. Circulation: Cardiovascular Genetics. 2010:CIRCGENETICS. 109.934455.
119. Palmer TM, Nordestgaard BG, Benn M, Tybjærg-Hansen A, Smith GD, Lawlor DA, et al. Association of plasma uric acid with ischaemic heart disease and blood pressure: mendelian randomisation analysis of two large cohorts. Bmj. 2013;347:f4262.
120. White J, Sofat R, Hemani G, Shah T, Engmann J, Dale C, et al. Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. The lancet Diabetes & endocrinology. 2016;4(4):327-36.
121. van Kippersluis H, O'Donnell O, van Doorslaer E. Long run returns to education: does schooling lead to an extended old age? Eur J Hum Resour. 2009;4:1-33.
122. Davies NM, Dickson M, Davey Smith G, van den Berg G, Windmeijer F. The Causal Effects of Education on Health, Mortality, Cognition, Well-being, and Income in the UK Biobank. bioRxiv. 2016.
123. Lleras-Muney A. The relationship between education and adult mortality in the United States. Review of Economic Studies. 2005;72:189-221.
124. Mazumder B. Does education improve health? A reexamniation of the evidence from compulsory schooling laws. Economic Perspectives. 2008;33:2-16.
125. Lager AC, Torssander J. Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(22):8461-6.

126. Madsen M, Andersen AM, Christensen K, Andersen PK, Osler M. Does educational status impact adult mortality in Denmark? A twin approach. Am J Epidemiol. 2010;172(2):225-34.
127. Madsen M, Andersen PK, Gerster M, Andersen AM, Christensen K, Osler M. Are the educational differences in incidence of cardiovascular disease explained by underlying familial factors? A twin study. Social science & medicine (1982). 2014;118:182-90.
128. Lundborg P. The health returns to schooling - what can we learn from twins? J Pop Economics. 2013;26:673-701.
129. Arden R, Luciano M, Deary IJ, Reynolds CA, Pedersen NL, Plassman BL, et al. The association between intelligence and lifespan is mostly genetic. Int J Epidemiol. 2016;45(1):178-85.
130. Lundborg P, Lyttkens CH, Nystedt P. The Effect of Schooling on Mortality: New Evidence From 50,000 Swedish Twins. Demography. 2016;53(4):1135-68.
131. Naess O, Hoff DA, Lawlor D, Mortensen LH. Education and adult cause-specific mortality--examining the impact of family factors shared by 871 367 Norwegian siblings. Int J Epidemiol. 2012;41(6):1683-91.
132. Sondergaard G, Mortensen LH, Nybo Andersen AM, Andersen PK, Dalton SO, Madsen M, et al. Does shared family background influence the impact of educational differences on early mortality? Am J Epidemiol. 2012;176(8):675-83.
133. McFarland MJ, Wagner BG. Does a college education reduce depressive symptoms in American young adults? Social Science & Medicine. 2015;146:75-84.
134. R E Marioni, S J Ritchie, P K Joshi, S P Hagenaars, A Okbay, K Fischer, et al. Genetic variants linked to education predict longevity. PNAS. 2016; in press.
135. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science. 2013;340(6139):1467-71.
136. Nguyen TT, Tchetgen Tchetgen EJ, Kawachi I, Gilman SE, Walter S, Liu SY, et al. Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk. Annals of epidemiology. 2016;26(1):71-6.e1-3.
137. Cuellar-Partida G, Lu Y, Kho PF, Hewitt AW, Wichmann HE, Yazar S, et al. Assessing the Genetic Predisposition of Education on Myopia: A Mendelian Randomization Study. Genetic epidemiology. 2016;40(1):66-72.
138. Case A, Fertig A, Paxson C. The lasting impact of childhood health and circumstance. Journal of health economics. 2005;24(2):365-89.
139. Smith JP. The impact of socioeconomic status on health over the life-course. Journal of Human Resources. 2007;42(4):739-64.
140. Voss M, Nylén L, Floderus B, Diderichsen F, Terry PD. Unemployment and early cause-specific mortality: a study based on the Swedish twin registry. American journal of public health. 2004;94(12):2155-61.
141. Martikainen PT, Valkonen T. Excess mortality of unemployed men and women during a period of rapidly increasing unemployment. The Lancet. 1996;348(9032):909-12.
142. Reeves A, Karanikolos M, Mackenbach J, McKee M, Stuckler D. Do employment protection policies reduce the relative disadvantage in the

labour market experienced by unhealthy people? A natural experiment created by the Great Recession in Europe. Social Science & Medicine. 2014;121:98-108.
143. Chang S-S, Stuckler D, Yip P, Gunnell D. Impact of 2008 global economic crisis on suicide: time trend study in 54 countries. Bmj. 2013;347:f5239.
144. Toffolutti V, Suhrcke M. Assessing the short term health impact of the Great Recession in the European Union: a cross-country panel analysis. Preventive medicine. 2014;64:54-62.
145. Margerison-Zilko C, Goldman-Mellor S, Falconi A, Downing J. Health Impacts of the Great Recession: a Critical Review. Current epidemiology reports. 2016;3(1):81-91.
146. Ruhm CJ. Recessions, healthy no more? Journal of health economics. 2015;42:17-28.
147. Stuckler D, Basu S, Suhrcke M, Coutts A, McKee M. The public health effect of economic crises and alternative policy responses in Europe: an empirical analysis. The Lancet. 2009;374(9686):315-23.
148. Noelke C, Avendano M. Who suffers during recessions? Economic downturns, job loss, and cardiovascular disease in older Americans. American journal of epidemiology. 2015;182(10):873-82.
149. Osler M, McGue M, Lund R, Christensen K. Marital status and twins' health and behavior: an analysis of middle-aged Danish twins. Psychosomatic medicine. 2008;70(4):482.
150. Gulliksson M, Burell G, Vessby B, Lundin L, Toss H, Svärdsudd K. Randomized controlled trial of cognitive behavioral therapy vs standard treatment to prevent recurrent cardiovascular events in patients with coronary heart disease: Secondary Prevention in Uppsala Primary Health Care project (SUPRIM). Archives of Internal Medicine. 2011;171(2):134-40.
151. Kuulasmaa K, Tunstall-Pedoe H, Dobson A, Fortmann S, Sans S, Tolonen H, et al. Estimation of contribution of changes in classic risk factors to trends in coronary-event rates across the WHO MONICA Project populations. The Lancet. 2000;355(9205):675-87.
152. Gasparrini A, Armstrong B, Kenward M. Multivariate meta-analysis for non-linear and other multi-parameter associations. Statistics in medicine. 2012;31(29):3821-39.
153. Bandosz P, O'Flaherty M, Drygas W, Rutkowski M, Koziarek J, Wyrzykowski B, et al. Decline in mortality from coronary heart disease in Poland after socioeconomic transformation: modelling study. bmj. 2012;344:d8136.
154. Bruthans J, Cífková R, Lánská V, O'flaherty M, Critchley JA, Holub J, et al. Explaining the decline in coronary heart disease mortality in the Czech Republic between 1985 and 2007. European journal of preventive cardiology. 2014;21(7):829-39.
155. Bobak M, Marmot M. East-West mortality divide and its potential explanations: proposed research agenda. The European Journal of General Practice. 1996;2(1):8-.
156. Ginter E. Cardiovascular risk factors in the former communist countries. European journal of epidemiology. 1995;11(2):199-205.
157. Kesteloot H, Sans S, Kromhout D. Dynamics of cardiovascular and all-cause mortality in Western and Eastern Europe between 1970 and 2000. European heart journal. 2005;27(1):107-13.

158. Pomerleau J, McKee M, Lobstein T, Knai C. The burden of disease attributable to nutrition in Europe. Public health nutrition. 2003;6(5):453-61.
159. Zatonski WA, McMichael AJ, Powles JW. Ecological study of reasons for sharp decline in mortality from ischaemic heart disease in Poland since 1991. Bmj. 1998;316(7137):1047.
160. Powles JW, Zatonski W, Vander Hoorn S, Ezzati M. The contribution of leading diseases and risk factors to excess losses of healthy life in Eastern Europe: burden of disease study. BMC public health. 2005;5(1):116.
161. Stefler D, Pikhart H, Jankovic N, Kubinova R, Pajak A, Malyutina S, et al. Healthy diet indicator and mortality in Eastern European populations: prospective evidence from the HAPIEE cohort. European Journal of Clinical Nutrition. 2014;68(12):1346.
162. Leon DA, Chenet L, Shkolnikov VM, Zakharov S, Shapiro J, Rakhmanova G, et al. Huge variation in Russian mortality rates 1984–94: artefact, alcohol, or what? The lancet. 1997;350(9075):383-8.
163. Britton A, McKee M. The relation between alcohol and cardiovascular disease in Eastern Europe: explaining the paradox. Journal of Epidemiology & Community Health. 2000;54(5):328-32.
164. McKee M, Britton A. The positive relationship between alcohol and heart disease in eastern Europe: potential physiological mechanisms. Journal of the Royal Society of Medicine. 1998;91(8):402-7.
165. Shkolnikov VM, Meslé F. The Russian epidemiological crisis as mirrored by mortality trends. Russia's Demographic Crisis, Conference Proceedings 1996.
166. Rehm J, Sulkowska U, Mańczuk M, Boffetta P, Powles J, Popova S, et al. Alcohol accounts for a high proportion of premature mortality in central and eastern Europe. International journal of epidemiology. 2007;36(2):458-67.
167. Leon DA, Saburova L, Tomkins S, Andreev E, Kiryanov N, McKee M, et al. Hazardous alcohol drinking and premature mortality in Russia: a population based case-control study. The Lancet. 2007;369(9578):2001-9.
168. Grittner U, Kuntsche S, Gmel G, Bloomfield K. Alcohol consumption and social inequality at the individual and country levels—results from an international study. The European Journal of Public Health. 2012;23(2):332-9.
169. Zaridze D, Brennan P, Boreham J, Boroda A, Karpov R, Lazarev A, et al. Alcohol and cause-specific mortality in Russia: a retrospective case–control study of 48 557 adult deaths. The Lancet. 2009;373(9682):2201-14.
170. Zaridze D, Lewington S, Boroda A, Scélo G, Karpov R, Lazarev A, et al. Alcohol and mortality in Russia: prospective observational study of 151 000 adults. The Lancet. 2014;383(9927):1465-73.
171. Mackenbach JP, Kulhánová I, Bopp M, Borrell C, Deboosere P, Kovács K, et al. Inequalities in alcohol-related mortality in 17 European countries: a retrospective analysis of mortality registers. PLoS medicine. 2015;12(12):e1001909.
172. Bobak M, Malyutina S, Horvat P, Pajak A, Tamosiunas A, Kubinova R, et al. Alcohol, drinking pattern and all-cause, cardiovascular and alcohol-related mortality in Eastern Europe. European Journal of Epidemiology. 2016;31(1):21-30.
173. Holmes MV, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, et al. Association between alcohol and cardiovascular disease: Mendelian

randomisation analysis based on individual participant data. Bmj. 2014;349:g4164.
174. The_World_Bank. International Comparison Program database. Accessed on 17 Nov 2017. 2017 [Available from: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD?locations=GR-RU-1W.
175. Marmot M, Bobak M. International comparators and poverty and health in Europe. BMJ. 2000;321(7269):1124.
176. Stuckler D, King L, McKee M. Mass privatisation and the post-communist mortality crisis: a cross-national analysis. The Lancet. 2009;373(9661):399-407.
177. Jukkala T, Stickley A, Mäkinen IH, Baburin A, Sparén P. Age, period and cohort effects on suicide mortality in Russia, 1956− 2005. BMC public health. 2017;17(1):235.
178. Feigin V. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016;388(10053):1459-544.
179. Sobotka T, Skirbekk V, Philipov D. Economic recession and fertility in the developed world. Population and development review. 2011;37(2):267-306.
180. Shkolnikov VM, Leon DA, Adamets S, Andreev E, Deev A. Educational level and adult mortality in Russia: an analysis of routine data 1979 to 1994. Social Science & Medicine. 1998;47(3):357-69.
181. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. The American journal of cardiology. 1976;38(1):46-51.
182. Brittain E. Probability of coronary heart disease developing. Western Journal of Medicine. 1982;136(1):86.
183. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. American heart journal. 1991;121(1):293-8.
184. Anderson KM, Wilson P, Odell PM, Kannel WB. An updated coronary risk profile. A statement for health professionals. Circulation. 1991;83(1):356-62.
185. Conroy R, Pyörälä K, Fitzgerald Ae, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. European heart journal. 2003;24(11):987-1003.
186. Woodward M, Brindle P, Tunstall-Pedoe H. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). Heart. 2007;93(2):172-6.
187. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. Bmj. 2007;335(7611):136.
188. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. Bmj. 2008;336(7659):1475-82.

189. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk. Circulation. 2014;129(25 suppl 2):S49-S73.
190. Ramachandran S, French J, Vanderpump M, Croft P, Neary R. Using the Framingham model to predict heart disease in the United Kingdom: retrospective study. Bmj. 2000;320(7236):676-7.
191. Ridker PM, Cook NR. Statins: new American guidelines for prevention of cardiovascular disease. The Lancet. 2013;382(9907):1762-5.
192. Cook NR, Ridker PM. Further insight into the cardiovascular risk calculator: the roles of statins, revascularizations, and underascertainment in the Women's Health Study. JAMA internal medicine. 2014;174(12):1964-71.
193. DeFilippis AP, Young R, Carrubba CJ, McEvoy JW, Budoff MJ, Blumenthal RS, et al. An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohortcalibration and discrimination among CVD risk scores. Annals of internal medicine. 2015;162(4):266-75.
194. Woodward M. On validation of cardiovascular risk scores. BMJ. 2016.
195. Brindle PM, McConnachie A, Upton MN, Hart CL, Smith GD, Watt GC. The accuracy of the Framingham risk-score in different socioeconomic groups: a prospective study. Br J Gen Pract. 2005;55(520):838-45.
196. Tunstall-Pedoe H, Woodward M. By neglecting deprivation, cardiovascular risk scoring will exacerbate social gradients in disease. Heart. 2006;92(3):307-10.
197. Organization WH. Prevention of cardiovascular disease: guidelines for assessment and management of cardiovascular risk: World Health Organization; 2007.
198. Hajifathalian K, Ueda P, Lu Y, Woodward M, Ahmadvand A, Aguilar-Salinas CA, et al. A novel risk score to predict cardiovascular disease risk in national populations (Globorisk): a pooled analysis of prospective cohorts and health examination surveys. The Lancet Diabetes & Endocrinology. 2015;3(5):339-55.
199. Ueda P, Woodward M, Lu Y, Hajifathalian K, Al-Wotayan R, Aguilar-Salinas CA, et al. Laboratory-based and office-based risk scores and charts to predict 10-year risk of cardiovascular disease in 182 countries: a pooled analysis of prospective cohorts and health surveys. The Lancet Diabetes & Endocrinology. 2017;5(3):196-213.
200. Gakidou E, Afshin A, Abajobir AA, Abate KH, Abbafati C, Abbas KM, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1345-422.
201. Collaboration APCS. A comparison of the associations between risk factors and cardiovascular disease in Asia and Australasia. European Journal of Cardiovascular Prevention & Rehabilitation. 2005;12(5):484-91.
202. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet (London, England). 2004;364(9438):937-52.
203. Jdanov DA, Deev AD, Jasilionis D, Shalnova SA, Shkolnikova MA, Shkolnikov VM. Recalibration of the SCORE risk chart for the Russian population. European journal of epidemiology. 2014;29(9):621-8.

204. Ruwanpathirana T, Owen A, Reid CM. Review on cardiovascular risk prediction. Cardiovascular therapeutics. 2015;33(2):62-70.
205. Sattar N, Welsh P, Sarwar N, Danesh J, Di Angelantonio E, Gudnason V, et al. NT-proBNP is associated with coronary heart disease risk in healthy older women but fails to enhance prediction beyond established risk factors: results from the British Women's Heart and Health Study. Atherosclerosis. 2010;209(1):295-9.
206. Polonsky TS, McClelland RL, Jorgensen NW, Bild DE, Burke GL, Guerci AD, et al. Coronary artery calcium score and risk classification for coronary heart disease prediction. Jama. 2010;303(16):1610-6.
207. Yeboah J, McClelland RL, Polonsky TS, Burke GL, Sibley CT, O'Leary D, et al. Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals. Jama. 2012;308(8):788-95.
208. Abraham G, Havulinna AS, Bhalala OG, Byars SG, De Livera AM, Yetukuri L, et al. Genomic prediction of coronary heart disease. European heart journal. 2016;37(43):3267-78.
209. Tada H, Melander O, Louie JZ, Catanese JJ, Rowland CM, Devlin JJ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. European heart journal. 2015;37(6):561-7.
210. Tikkanen E, Havulinna AS, Palotie A, Salomaa V, Ripatti S. Genetic Risk Prediction and a 2-Stage Risk Screening Strategy for Coronary Heart DiseaseSignificance. Arteriosclerosis, Thrombosis, and Vascular Biology. 2013;33(9):2261-6.
211. Mega JL, Stitziel NO, Smith JG, Chasman DI, Caulfield MJ, Devlin JJ, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. The Lancet. 2015;385(9984):2264-71.
212. Ingle L, Carroll S, Stamatakis E, Hamer M. Benefit of adding lifestyle-related risk factors for prediction of cardiovascular death among cardiac patients. International journal of cardiology. 2013;163(2):196-200.
213. Georgousopoulou EN, Panagiotakos DB, Bougatsas D, Chatzigeorgiou M, Kavouras SA, Chrysohoou C, et al. Physical Activity Level Improves the Predictive Accuracy of Cardiovascular Disease Risk Score: The ATTICA Study (2002–2012). International journal of preventive medicine. 2016;7.
214. Graversen P, Abildstrøm SZ, Jespersen L, Borglykke A, Prescott E. Cardiovascular risk prediction: Can Systematic Coronary Risk Evaluation (SCORE) be improved by adding simple risk markers? Results from the Copenhagen City Heart Study. European journal of preventive cardiology. 2016;23(14):1546-56.
215. Sacco RL, Khatri M, Rundek T, Xu Q, Gardener H, Boden-Albala B, et al. Improving global vascular risk prediction with behavioral and anthropometric factors: the multiethnic NOMAS (Northern Manhattan Cohort Study). Journal of the American College of Cardiology. 2009;54(24):2303-11.
216. Paynter NP, LaMonte MJ, Manson JE, Martin LW, Phillips L, Ridker PM, et al. Comparison of Lifestyle Based to Traditional CVD Prediction in a Multiethnic Cohort of Non-Smoking Women. Circulation. 2014:CIRCULATIONAHA. 114.012069.

217. Fiscella K, Tancredi D, Franks P. Adding socioeconomic status to Framingham scoring to reduce disparities in coronary risk assessment. American heart journal. 2009;157(6):988-94.
218. Ramsay SE, Morris RW, Whincup PH, Papacosta AO, Thomas MC, Wannamethee SG. Prediction of coronary heart disease risk by Framingham and SCORE risk assessments varies by socioeconomic position: results from a study in British men. European Journal of Cardiovascular Prevention & Rehabilitation. 2011;18(2):186-93.
219. Pujades-Rodriguez M, Timmis A, Stogiannis D, Rapsomaniki E, Denaxas S, Shah A, et al. Socioeconomic deprivation and the incidence of 12 cardiovascular diseases in 1.9 million women and men: implications for risk prediction and prevention. PloS one. 2014;9(8):e104671.
220. Veronesi G, Gianfagna F, Giampaoli S, Chambless L, Mancia G, Cesana G, et al. Improving long-term prediction of first cardiovascular event: the contribution of family history of coronary heart disease and social status. Preventive medicine. 2014;64:75-80.
221. Colantonio LD, Richman JS, Carson AP, Lloyd-Jones DM, Howard G, Deng L, et al. Performance of the Atherosclerotic Cardiovascular Disease Pooled Cohort Risk Equations by Social Deprivation Status. Journal of the American Heart Association. 2017;6(3):e005676.
222. Backholer K, Hirakawa Y, Tonkin A, Giles G, Magliano DJ, Colagiuri S, et al. Development of an Australian cardiovascular disease mortality risk score using multiple imputation and recalibration from national statistics. BMC cardiovascular disorders. 2017;17(1):17.
223. Havranek EP, Mujahid MS, Barr DA, Blair IV, Cohen MS, Cruz-Flores S, et al. Social determinants of risk and outcomes for cardiovascular disease. Circulation. 2015;132(9):873-98.
224. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22.
225. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ. 2005;330(7499):1076-9.
226. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.
227. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology. 2015;44:512-25.
228. Burgess S, Dudbridge F, Thompson SG. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". Am J Epidemiol. 2015;181:290-1.
229. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet. 2012;379(9822):1214-24.
230. Holmes MV, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. BMJ : British Medical Journal. 2014;349.

231. White J, Swerdlow DI, Preiss D, et al. Association of lipid fractions with risks for coronary artery disease and diabetes. JAMA Cardiology. 2016;1(6):692-9.
232. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2016.
233. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nature genetics. 2015;47(11):1236-41.
234. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539-42.
235. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016. doi: 10.1002/gepi.21998. [Epub ahead of print].
236. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. International journal of epidemiology. 2013;42(5):1497-501.
237. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nature genetics. 2013;45(1):25-33.
238. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. International journal of epidemiology. 2016.
239. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomisation via the zero modal pleiotropy assumption. International Journal of Epidemiology. 2017;[forthcoming].
240. Greco M FD, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Statistics in Medicine. 2015;34(21):2926-40.
241. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nature genetics. 2011;43(4):333-8.
242. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nature genetics. 2011;43(4):339-44.
243. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature genetics. 2015;47(10):1121-30.
244. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. American journal of epidemiology. 2015;181(4):251-60.
245. Peasey A, Bobak M, Kubinova R, Malyutina S, Pajak A, Tamosiunas A, et al. Determinants of cardiovascular disease and other non-communicable diseases in Central and Eastern Europe: rationale and design of the HAPIEE study. BMC public health. 2006;6:255.

246. Evans A, Salomaa V, Kulathinal S, Asplund K, Cambien F, Ferrario M, et al. MORGAM (an international pooling of cardiovascular cohorts). Int J Epidemiol. 2005;34(1):21-7.
247. Bobak M, Pikhart H, Hertzman C, Rose R, Marmot M. Socioeconomic factors, perceived control and self-reported health in Russia. A cross-sectional survey. Social Science & Medicine. 1998;47(2):269-79.
248. Collaboration PS. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. The Lancet. 2002;360(9349):1903-13.
249. Rehm J. Measuring quantity, frequency, and volume of drinking. Alcoholism: clinical and experimental research. 1998;22(S2).
250. Lane PW, Nelder JA. Analysis of covariance and standardization as instances of prediction. Biometrics. 1982:613-21.
251. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int J Epidemiol. 2015;44(4):1137-47.
252. Fischer K, Kettunen J, Wurtz P, Haller T, Havulinna AS, Kangas AJ, et al. Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. PLoS medicine. 2014;11(2):e1001606.
253. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). European Heart Journal. 2016;37(29):2315-81.
254. Kivimaki M, Ferrie JE, Batty GD, Davey Smith G, Elovainio M, Marmot MG, et al. Optimal form of operationalizing BMI in relation to all-cause and cause-specific mortality: the original Whitehall study. Obesity (Silver Spring). 2008;16(8):1926-32.
255. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Medical Decision Making. 2006;26(6):565-74.
256. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. The Lancet. 2016;388(10059):2532-61.
257. Hingorani A, Humphries S. Nature's randomised trials. The Lancet.366(9501):1906-8.
258. Madsen M, Andersen PK, Gerster M, Andersen A-MN, Christensen K, Osler M. Are the educational differences in incidence of cardiovascular disease explained by underlying familial factors? A twin study. Social Science & Medicine. 2014;118:182-90.
259. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. International Journal of Epidemiology. 2016;45(6):1866-86.
260. Hansen KT, Heckman JJ, Mullen KJ. The effect of schooling and ability on achievement test scores. Journal of econometrics. 2004;121(1-2):39-98.
261. de Walque D. Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education. Journal of health economics. 2007;26(5):877-95.

262. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JE, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. Lancet. 2015;385(9965):351-61.
263. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. J Am Coll Cardiol. 2012;60(25):2631-9.
264. Fletcher JM. The promise and pitfalls of combining genetic and economic research. Health economics. 2011;20(8):889-92.
265. Kvaavik E, Glymour M, Klepp KI, Tell GS, Batty GD. Parental education as a predictor of offspring behavioural and physiological cardiovascular disease risk factors. European journal of public health. 2012;22(4):544-50.
266. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimaki M, et al. Association of socioeconomic position with health behaviors and mortality. Jama. 2010;303(12):1159-66.
267. Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. Hum Mol Genet. 2016;25(R2):R149-r56.
268. Varbo A, Benn M, Smith GD, Timpson NJ, Tybjaerg-Hansen A, Nordestgaard BG. Remnant cholesterol, low-density lipoprotein cholesterol, and blood pressure as mediators from obesity to ischemic heart disease. Circulation research. 2015;116(4):665-73.
269. Botero J, Ponce A, Shleifer A. Education, Complaints, and Accountability. J Law Econ. 2013;56(4):959-96.
270. Oreopoulos P, Salvanes KG. Priceless: The Nonpecuniary Benefits of Schooling. J Econ Perspect. 2011;25(1):159-84.
271. Filakti H, Fox J. Differences in mortality by housing tenure and by car access from the OPCS Longitudinal Study. Population Trends. 1995(81):27-30.
272. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimaki M, et al. Association of socioeconomic position with health behaviors and mortality. Jama. 2010;303(12):1159-66.
273. Kershaw KN, Droomers M, Robinson WR, Carnethon MR, Daviglus ML, Verschuren WM. Quantifying the contributions of behavioral and biological risk factors to socioeconomic disparities in coronary heart disease incidence: the MORGEN study. European Journal of Epidemiology. 2013;28(10):807-14.
274. Marmot MG, Bosma H, Hemingway H, Brunner E, Stansfeld S. Contribution of job control and other risk factors to social variations in coronary heart disease incidence. The Lancet. 1997;350(9073):235-9.
275. Kivimäki M, Nyberg ST, Batty GD, Fransson EI, Heikkilä K, Alfredsson L, et al. Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data. The Lancet. 2012;380(9852):1491-7.
276. Fiscella K, Franks P. Does psychological distress contribute to racial and socioeconomic disparities in mortality? Social Science & Medicine. 1997;45(12):1805-9.
277. Hintsa T, Shipley MJ, Gimeno D, Elovainio M, Chandola T, Jokela M, et al. Do pre-employment influences explain the association between

psychosocial factors at work and coronary heart disease? The Whitehall II study. Occupational and Environmental Medicine. 2010;67(5):330-4.
278. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. American journal of epidemiology. 2017;186(9):1026-34.
279. Vikhireva O, Pająk A, Broda G, Malyutina S, Tamosiunas A, Kubinova R, et al. SCORE performance in Central and Eastern Europe and former Soviet Union: MONICA and HAPIEE results. European heart journal. 2013;35(9):571-7.
280. Gaziano TA, Young CR, Fitzmaurice G, Atwood S, Gaziano JM. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-up Study cohort. The Lancet. 2008;371(9616):923-31.
281. Gaziano TA, Abrahams-Gessel S, Denman CA, Montano CM, Khanam M, Puoane T, et al. An assessment of community health workers' ability to screen for cardiovascular disease risk with a simple, non-invasive risk assessment instrument in Bangladesh, Guatemala, Mexico, and South Africa: an observational study. The Lancet Global Health. 2015;3(9):e556-e63.
282. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for individual Prognosis or diagnosis (TRIPOD): Explanation and Elaboration. Annals of internal medicine. 2015;162(1):W1-W73.
283. Harrell F. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed: Springer; 2015.
284. Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating: Springer Science & Business Media; 2008.
285. Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, et al. How to develop a more accurate risk prediction model when there are few events. Bmj. 2015;351:h3868.
286. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. 1996;15(4):361-87.
287. Peirce CS. The numerical measure of the success of predictions. Science. 1884;4(93):453-4.
288. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net Reclassification Improvement: Computation, Interpretation, and ControversiesA Literature Review and Clinician's Guide. Annals of internal medicine. 2014;160(2):122-31.
289. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk-prediction instruments: a critical review. Epidemiology (Cambridge, Mass). 2014;25(1):114.

Annex 1 – Questionnaire items (HAPIEE study)

A1.1. Socioeconomic factors

Education

3. What is your highest completed level of education?

1. Incomplete primary or no formal education
2. Primary
3. Vocational (apprenticeship)
4. Secondary
5. University (degree)

Unemployment (current and previous)

80. What is your current economic activity?

1. Employed
2. Entrepreneur (owner of a company)
3. Self-employed / freelance
4. Housewife
5. Farmer
6. Pensioner, still employed
7. Pensioner, not employed. At what age did you retire ? …….. ☐☐ years old
8. Unemployed

82. Have you ever experienced unemployment?

1. No
2. Yes, for up to 3 months in total
3. Yes, for 3 months to 1 year
4. Yes, for more than one year

Material depravation (current, and early life)

71. How often does it happen that you do not have enough **money for food** which you and your family need? And how often did this happen before 1990? (Russia = enough money to buy food)

at present	before 1990
☐ 1. all the time	☐ 1. all the time
☐ 2. often	☐ 2. often
☐ 3. sometimes	☐ 3. sometimes
☐ 4. rarely	☐ 4. rarely
☐ 5. never	☐ 5. never

72. How often does it happen that you do not have enough **money for clothing** which you and your family need? And how often did this happen before 1990?

at present	before 1990
☐ 1. all the time	☐ 1. all the time
☐ 2. often	☐ 2. often
☐ 3. sometimes	☐ 3. sometimes
☐ 4. rarely	☐ 4. rarely
☐ 5. never	☐ 5. never

73. Do you have difficulties with **paying bills** (for housing, electricity, heating etc)? And what was the situation before 1990?

at present	before 1990
☐ 1. all the time	☐ 1. all the time
☐ 2. often	☐ 2. often
☐ 3. sometimes	☐ 3. sometimes
☐ 4. rarely	☐ 4. rarely
☐ 5. never	☐ 5. never

Material amenities (current and early life)

84. Now, would you tell us about your household? Below is a list of various items, which of the following do you have in your household?

	Yes	No, I do not want it	No, I can not afford it
Microwave	☐ 1	☐ 2	☐ 3
Video recorder	☐ 1	☐ 2	☐ 3
Television (colour)	☐ 1	☐ 2	☐ 3
Washing machine	☐ 1	☐ 2	☐ 3
Dishwasher	☐ 1	☐ 2	☐ 3
Car	☐ 1	☐ 2	☐ 3
Freezer	☐ 1	☐ 2	☐ 3
Cottage (for holidays / weekends etc.)	☐ 1	☐ 2	☐ 3
Videocamera / camcorder	☐ 1	☐ 2	☐ 3
Satellite / cable TV	☐ 1	☐ 2	☐ 3
Telephone	☐ 1	☐ 2	☐ 3
Mobile phone	☐ 1	☐ 2	☐ 3

79. Did you have any of the following items in your house when you were a child (about 10 years old)?

Cold tap water	1. Yes	2. No	3. I don't remember
Hot tap water	1. Yes	2. No	3. I don't remember
Radio	1. Yes	2. No	3. I don't remember
Fridge	1. Yes	2. No	3. I don't remember
Own kitchen	1. Yes	2. No	3. I don't remember
Own toilet	1. Yes	2. No	3. I don't remember

A1.2. Psychosocial factors

Depression

61. Below is a list of the ways you might have felt or behaved during **the last week**.
For each of the following statements, please indicate how often you felt that way:

During the past week:	*Less than one day*	*1-2 days*	*3-4 days*	*5-7 days*
a) I was bothered by things that usually do not bother me	☐ 1	☐ 2	☐ 3	☐ 4
b) I did not feel like eating, my appetite was poor	☐ 1	☐ 2	☐ 3	☐ 4
c) I felt that I could not shake off the blues even with help from my family and friends	☐ 1	☐ 2	☐ 3	☐ 4
d) I felt that I was just as good as other people	☐ 1	☐ 2	☐ 3	☐ 4
e) I had trouble keeping my mind on what I was doing	☐ 1	☐ 2	☐ 3	☐ 4
f) I felt depressed.	☐ 1	☐ 2	☐ 3	☐ 4
g) I felt that everything I did was an effort	☐ 1	☐ 2	☐ 3	☐ 4
h) I felt hopeful about the future	☐ 1	☐ 2	☐ 3	☐ 4
i) I thought my life had been a failure	☐ 1	☐ 2	☐ 3	☐ 4
j) I felt fearful	☐ 1	☐ 2	☐ 3	☐ 4
k) My sleep was restless	☐ 1	☐ 2	☐ 3	☐ 4
l) I was happy	☐ 1	☐ 2	☐ 3	☐ 4
m)I talked less than usual	☐ 1	☐ 2	☐ 3	☐ 4
n) I felt lonely	☐ 1	☐ 2	☐ 3	☐ 4
o) People were unfriendly	☐ 1	☐ 2	☐ 3	☐ 4
p) I enjoyed life	☐ 1	☐ 2	☐ 3	☐ 4
q) I had crying spells	☐ 1	☐ 2	☐ 3	☐ 4
r) I felt sad	☐ 1	☐ 2	☐ 3	☐ 4
s) I felt people dislike me	☐ 1	☐ 2	☐ 3	☐ 4
t) I could not get going	☐ 1	☐ 2	☐ 3	☐ 4

Perceived control

70. How much do you agree or disagree with the following statements?

	DISAGREE			AGREE		
	STRONGLY	MODERATELY	SLIGHTLY	SLIGHTLY	MODERATELY	STRONGLY
a) At home, I feel I have control over what happens in most situations	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b) Keeping healthy depends on things that I can do	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
c) There are certain things I can do for myself to reduce the risk of a heart attack	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
d) There are certain things I can do for myself to reduce the risk of getting cancer	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
e) I feel that what happens in my life is often determined by factors beyond my control	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
f) Over the next 5-10 years I expect to have many more positive than negative experiences	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
g) I often have the feeling that I am being treated unfairly	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
h) In the past ten years my life has been full of changes without my knowing what will happen next	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
i) I very often have the feeling that there's little meaning in the things I do in my daily life	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
j) I sometimes feel as if I've done all there is to do in life	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
k) I gave up trying to make big improvements or changes in my life a long time ago	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

Social change

69. Have the changes since 1989 been good or bad for you:

	Very good	*Good*	*No change*	*Bad*	*Very bad*
Occupational position	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Income	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Material circumstances	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
General social position	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Social Support

4. What is your marital status?

1. Single
2. Married
3. Cohabiting
4. Russia: Divorced / Separated
5. Widowed

64. Are you regularly in **contact with your relatives** who do not live in your household?

1. several times a week
2. about once a week
3. several times a month
4. about once a month
5. less than once a month
6. I do not have relatives / no relatives outside my household

66. How often do you visit **friends**?

1. several times a week
2. about once a week
3. several times a month
4. about once a month
5. less than once a month
6. I do not have friends

62. Are you **a member of club or organisation** (sports club, church, political party)?

1. Yes
2. No

If **YES**, how often do you take part in common activities?

1. Several times a week
2. Several times a month
3. About once a month
4. Several times a year
5. Never or almost never

Annex 2 – Best practice in model development and evaluation

In writing Annex 2, I have mainly been influenced by the 2015 Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Statement,(282) which was the first attempt to agree and disseminate common standards of best practice to a wide platform of the epidemiological and medical research community. In some instances where the TRIPOD statement does not go into sufficient detail, I have added additional guidance from two textbooks, written by two of the leading experts in this area of methodology, one written by Harrell,(283) and the other by Steyerberg.(284) Since this body of recommendations is remarkably consistent with each other (for example, I have never found an instance of disagreement between these three sources), I will forego detailed annotation as to which of the three sources supports each of the summary statements I make. I focus on those suggestions around model development which directly inform on the strengths and weaknesses of the existing models I reviewed in section 1.4.3, as well as methodological decisions I have to make around my own model development in section 2.3.

A2.1. Best practice in model development

A2.1.1. Cleaning and preparing the derivation dataset.

If >10% of data is **missing** for a key predictor of the outcome, then multiple imputation of these values will lead to models that are more valid (whereby validity was evaluated in studies where the true events were masked during model development, or in studies were validity was examined by testing performance in an external and independent dataset).

Continuous variables should not be categorized or dichotomized. If their distributions are not normal, then reasonable attempts should be made to normalize these (e.g. log transformation or similar). This is particularly important in the case of continuous variables that are collinear with each other, where one variable is left-skewed while the other is right-skewed (e.g. metabolites determined via proton nuclear magnetic resonance spectroscopy). After normalizing, the number of independent predictors can be lowered, while failure to do so may result in more seemingly independent predictors, whereby the extra predictors are merely an artefact from incomplete normalization. Too many predictors increase the risk of **overfitting**, that is models may be excessively tailored to the data used to derive the model, leading to artificially inflated measures of performance during model derivation. However, this will not replicate in external settings, where performance will drop proportionately to the degree of overfitting.

If continuous variables have nonlinear properties, then restricted cubic **splines** are the most effective way to fit these shapes (as opposed to choosing a-priori thresholds or fitting simple quadratic functions with a user-defined nadir of risk). Each knot in the spline will add another parameter, so an *a priori* number of knots should be specified. However, if the model is made with splines and the parameters need to be communicated to other researchers (for example, for external validation), then it can be difficult to extract and write out the spline function in written text format in the manuscript text. One solution is to supply the model in

electronic format (e.g. by creating a purpose-build website using the R package “shiny”). A reasonable compromise towards simplification is to revert to **quadratics** functions, since this can often fit biological data almost as well as splines, but are simpler to communicate, understand, and use in external settings.

A2.1.2. Variable selection

When deciding which candidate predictors to keep or **drop** for the final model, then the worst approach is to use **univariate tests of association** with the outcome. Consider if strong predictor A is associated strongly with bad predictor B, and where strong predictor A is associated weakly with good predictor C. Univariate testing will incorrectly identify A and subsequently B as the top predictors, while in a mutually adjusted model, A and C would be better predictors. Instead, predictor selection should be done *a priori* where possible and failing that, by an automated multivariable method. I will use the term “predictor selection”, which is thought to be synonymous with “variable selection” and “parameter selection” in statistical terminology, which is synonymous with “feature selection” in machine-learning terminology.

A priori approaches can combine considerations around:

1) Review of published **literature** for predictors that have demonstrated external validity before.

2) Clinical reasoning and **expert** opinion.
    a. Note that **causality** is an advantage, as causal predictors are more likely to replicate externally. However, causality is not a requirement, since noncausal predictors can sometimes perform better. For example, a causal predictor A can be displaced by a noncausal predictor B, in case that noncausal predictor B has a lower measurement error and thereby larger magnitude of association with the outcome of interest.

b. Another consideration is the **cultural preferences of clinicians and patients** who may be end users of the model. For example, some users may consider psychosocial factors are perceived as less legitimate (for example, staff trained in surgery or A+E) when compared to settings where psychosocial factors are more established (for example, paediatrics, primary care, geriatric medicine and palliative medicine). In part, this need not reflect the empirical evidence of utility, but also the personality types who are attracted to these respective specialties.

3) Perceived **collinearity** and overlap with more established predictors.

* = *I have not previously seen the following suggestions in the literature, but from my own reasoning, these seem natural extensions to this set of considerations.*

4) *Data **availability in validation** dataset(s). This is a pragmatic step for researchers, but (similar to the derivation of QRISK) will bias the final models towards including predictors that are already commonly established and accepted, as opposed to inducing potentially true predictors that have been less studied.

5) *Externalities created during data collection:
   a. Negative externalities – e.g. **cost**, discomfort, pain and/or **convenience** of variable collection (especially to increase response rate, and to maintain a positive long-term clinical relationship in otherwise asymptomatic and happy participants).
   b. Positive externalities – e.g. after collecting ‘omics data from the patient, to complete a particular clinical model, then the same ‘omics data may subsequently be **reused** for other utilities, outside this particular model and outcome.

i. For example, predictors that are relatively distal and upstream to the disease process (e.g. genomics, and socioeconomic and psychosocial variables) may have lower predictive power for a given outcome (such as CVD) when compared to a more proximal predictor (such as Troponin), however investments in the data infrastructure for upstream predictors are more likely to be repurposed to predict **other outcomes** (such as predicting cancer).
ii. Another secondary utility is not risk prediction but established **clinical** care. For example, genomic data is stable throughout the lifetime of the patient, and can be repurposed in case that the participant is later diagnosed with a malignancy with known gene-specific therapy (e.g. HER2-positive breast cancer).

Once the five *a priori* considerations have been exhausted, then *automatic predictor-selection* could be considered next, particularly if there are fewer than 10 to 20 events per each candidate parameter in the derivation dataset, to reduce the risk of overfitting. There remains disagreement around when to use each of the following, so much of the below reflects my personal understanding.

1) **Backwards-selection**: this remains the default approach, when it is desirable to eliminate a small proportion of variables, and where the candidate predictors have lower levels of collinearity. There are three ways of choosing between two competing models:
   a. Take the larger model, and inspect the weakest predictor in that model for its P-value. Predictors included in the final chosen model do not have to demonstrate independent associations with the outcome at conventional levels of significance (e.g. $P < 0.05$, whether adjusted for Bonferroni or not). Hence the optimum P-value threshold is typically relaxed to $0.1 < P < 0.2$.

b. Run a Likelihood-Ratio test comparing model A with model B. This is likely to result in a number comparable to the P-value of the weakest predictor. This approach is agnostic to the question of how overfitted the data are likely to be.
c. Compare the AIC of model A against model B as the criterion. This approach makes considerations for how overfitted the data are likely to be.

2) **Forwards-selection**: this is similar to backwards selection, and is preferred in cases of increasing collinearity between the candidate predictors (e.g. ‘omics data)

In cases when there are more predictors than events:(285)

3) and where there is a *small group of strong predictors* (i.e. most candidate predictors need to be dropped), then the **LASSO** (also known as L1 penalization technique) is increasingly recommended, since it shrinks many predictor coefficients to zero, thereby dropping these predictors.

4) alternatively, in case where a *large group of weak predictors* are thought to be superior, then **ridge regression** is recommended. This does not necessarily drop predictors, as beta coefficients are shrunken down *towards* zero but not necessarily *to* zero.

Of note, all these four methods of automatic predictor-selection will increase the possibility of overfitting the model (especially the forwards and backwards methods that give less consideration to the event-to-parameter ratio). Accordingly, when evaluating model performance:

a. if performing *internal* validation, then allowance for overfitting during predictor selection needs to be made, by incorporating predictor-selection into each bootstrapped sample.
b. If performing *external* validation, then allowance for overfitting during predictor-selection does not need to be specially accounted for, as this will already be detected in the external valuation sample.

Finally, the more sophisticated types of automatic predictor selection (such as LASSO and ridge) are beginning to resemble the methods of machine learning such as support vector machines (where this process is called “feature selection”). The main disadvantage of these methods is that, being nonparametric, they do not produce any measures of uncertainty, confidence intervals or p-values. They also require much more data to result in stable models. However, if uncertainty is less important and much data is available, then they can be exploited to fit more complex models. For example, many fit non-linear dose-response curves, as well as lots of interaction variables. However, nonparametric machine learning is beyond the scope of this review, so I will now return to parametric statistical methods.

A2.1.4. Interactions

Interactive effects can be specified *a priori* (if there are strong reasons to expect their occurrence from the previous literature or expert opinion). If these are absent, then Harrell advises against the process of screening individual interactions. Instead, he recommends testing a model with no interactions against a full model that incorporates a wide set of putative interactions. A single Log-Rank (LR) test can assess whether this bundle improves performance. However, it remains unclear to me what should be done if there is strong statistical support for the model with many interactions.

The practical application of interaction bundles is clearer in case that the interacting variable is binary. For example, a common historical practice

has been to split the model into multiple underlying models (for example, separate models based on gender or country, as was frequently done in the early CVD prediction models like SCORE). However, when should two separate models be fitted for men and women, and when should a single pooled model be created for them both, instead? My understanding is that a pooled model is always preferred, unless there is convincing reason to suggest otherwise. For example, model A could be a simple pooled model, and model B could be the same pooled model, which additionally includes all the possible gender-by-risk factor combinations. If an LR test finds strong statistical evidence that B is superior, then this would argue in favour of smaller, gender-specific models.

A2.1.5. Reducing overfit

The problem of deriving overfitted models remains a large one which is difficult to comprehensively mitigate. Many of the steps outlined in section A2.1. help towards this, however there are some specific steps that can be further done to reduce the problem of overfitted models, in some special cases.

Namely, if:

1) the final set of predictors has been chosen in a method that does not include shrinkage (i.e. the predictors were chosen by *a priori* reasoning, forwards- regression and/or stepwise-regression) AND
2) if the derivation dataset has <10 events per parameter, then

there remains a substantial risk that the model has been overfitted. This should be corrected with **shrinkage** or other coefficient penalization procedures. These are similar to the method of **ridge regression**: predictors are not dropped from the model, but their coefficients are merely shrunken.

A2.2. Best practice in evaluating model performance

Once the model has been developed, it is important to measure its performance as robustly as possibly, to give an indication of how the model might perform if used by other researchers or clinicians working on similar questions elsewhere. A key objective is to reliably capture, measure and account for any optimistic bias which may have resulted in an overfitted model.

A2.2.1. Selecting the dataset where performance is measured

Simpler approaches are listed at the top of this list, which moves on to increasingly superior approaches at the bottom of the list:

a. Performance can be evaluated in the derivation sample. This is never recommended due to the near certain probability of overfitting to a smaller or greater degree.

b. Performance can also be evaluated by dividing the initial dataset into two ***random* parts**, one for derivation and one for validation (perhaps with a 66:33 ratio of these two). Although used in the derivation of models like QRISK, this is now less advised, since omitting 33% of the data results in a less precise and a more overfitted model. This is a type of internal validation.

c. Divide the initial dataset into two ***non-random* parts**, either by using a geographical or temporal split. This is superior to method (b), since the validation data is less similar to derivation data, thereby reducing the possibility that spurious overfit will falsely be detected as “good performance” during validation. However, the problem of wasting data to derive more precise beta coefficients remains for both (b) and (c). This too is a type of internal validation.

d. Keep all the data together for model derivation. Generate **bootstrapped** datasets, in order to estimate **internal validity** (indeed,

this is what is most commonly meant by internal validity at time of writing). This approach carries forward the weakness of (b), since derivation and validation data are geographically more similar (in each bootstrap), thereby being less sensitive to detect any potential overfit. However, it is superior to both (b) and (c), since a larger derivation sample results in smaller measurement error in the beta coefficients used in the final model.

e. Perform model derivation in one independent dataset, and model validation in another **independent, external dataset**. The more different the two datasets are the better. This can mean differences in the country, sampling frame, inclusion criteria, response rate, data collection purpose (e.g. RCT, cohort, registry, electronic healthcare records), and outcome definition. This approach avoids the weaknesses of the internal validation approaches outlined in (b), (c), and (d), as the accidental overfit patterns in derivation are unlikely to also recapitulate in the external dataset.

A2.2.2. Introduction to prediction metrics

Unlike in other domains of epidemiology, where key results can often be interpreted in the light of a single statistical metric, risk prediction science is informed by holistic evaluation of multiple, mutually independent metrics. When evaluating the performance of a single model, and where this model returns predicted risks as a *continuous* outcome, then *Calibration* and *Discrimination* are the two more pertinent areas to assess. In case that predicted risks are subsequently dichotomised into categorical outcomes (which can assist in the implementation and interpretation of prediction models in clinical and public health settings, where interventional decisions are dichotomous in their nature), then *classification statistics* can additionally be used to complement *calibration* and *discrimination*.

When additionally looking at performance differences between two models, then measures of *Reclassification* can additionally help. These can handle both continuous and dichotomous risk prediction models.

I consider other measures that are sometimes used in risk prediction science (such as R^2 and the *Brier Score*) to add little additional information, as this is already reasonably reflected in the more clinically-established metrics of *calibration, discrimination, classification,* and *reclassification*.

A2.2.3. Single model - Calibration

Consider a hypothetical example, where a model predicts that 10% of the entire sample will develop an event. If the observed data also reports that 10% of the sample developed events, then this is known as perfect calibration. Some papers will rightly report how the Predicted/Observed ratio (P/O ratio) is 1.0. However, those precise individuals who were designated in the "at risk" group need not be those individuals who later developed the real life outcome. As thus, a perfectly calibrated model may theoretically be able to label as many as 100% of their predictions as *false positives*, if this is offset by an equal number of *false negatives*. Such a hypothetical example may be clinically useless, to advise doctors about whether they ought to prescribe the particular patient in front of them a statin. However, they are nonetheless very useful to those who are trying to plan and budget future healthcare resource needs (e.g. pharmaceutical companies, regional or national healthcare insurance system managers, such as NHS England) as they will be correctly provided with an aggregate measure of demand, calculated from a correct estimate of population health need. This example underscores the notion that calibration and discrimination are conceptually distinct measures. Good calibration is necessary and sufficient for population-level interventions. Good calibration is necessary but not sufficient for individual-level interventions.

In many tests of calibration, the population is broken down into smaller subgroups, so that calibration can be assessed within each subgroup. One approach is to break the sample into deciles or quintiles of risk. This allocates the same number of participants into each group. However, this means that there will often be many times fewer events in the subgroup with lowest risk, when compared to the subgroup with highest risk. It makes more sense to me, that subgroups could also be defined such that these contain equal or similar number of *events* at the end of the study, to further improve statistical robustness. I am not aware of discussion of this in the statistical literature. A third approach would be to define subgroups based on clinically meaningful thresholds of intervention. For example, when predicting CVD mortality, it might be useful to use the absolute risk thresholds 1%, 5% and 10% to create four categories, based on the 2016 European Society of Cardiology guidelines.(253) Despite seeming pragmatic, this third approach has been little discussed in the literature, and is rarely used in research papers reporting the results of calibration.

Once thresholds have been defined with which to divide the entire cohort into subgroups, then calibration needs to be evaluated. Visual calibration plots are the most common form. These are sometimes accompanied by formal statistical tests such as the Hosmer–Lemeshow test. However, this is typically too insensitive to detect poor calibration, and its result is heavily influenced by the sample size of the study. For this reason, most guidelines suggest that calibration be evaluated visually, where one dot represents one subgroup, x-axis denotes predicted risk in that subgroup (from 0-100%) and y-axis the observed risk in that subgroup (from 0-100%). These dots could either be left as they are for the eye to judge alone (whereby a 45-degree diagonal line denotes perfect calibration in all subgroups). Alternatively the dots could be connected with a straight line of best fit (permitting the calculation of the *calibration slope*); or alternatively connected with a smoothed (lowess) curve.

It helps to note that, at the level of the entire sample (where subgroups are not defined), calibration in the *derivation sample* may well be perfect

or near-perfect. This can be thought of as the result of a type of overfit that occurs in both small and large studies alike. For this reason, it is less useful to report overall calibration in the derivation sample, while calibration in the *validation* sample is more informative. In cases where validation data are not available, bootstrapping methods can be used to evaluate the degree of overfit in the derivation sample and correct for it.

A2.2.4. Single model - Discrimination

I now move onto measures of performance that are inherently individualistic in nature (as opposed to the group-based measure of calibration). These are useful if the intervention being considered is allocated to discrete individuals, by discrete clinicians (e.g. statin prescribing by GPs). Discrimination evaluates the probability that all cases (true positives plus false negatives) are assigned a higher predicted risk score when compared to all noncases (true negatives plus false positives). In logistic regression models (where follow-up time is not considered) discrimination is evaluated using the *Area Under the Receiver Operating Characteristic* (AUROC). In survival models (where follow-up time is considered, as is the case of most contemporary CVD risk prediction models), discrimination is evaluated with a C-statistic, such as Harrell's C-Statistic.(286)

As a summary, calibration is usually a function of how well the baseline risk of disease has been modelled (for participants with no risk factors, or average risk factors), while discrimination is usually a function of the Hazard Ratios ascribed to risk factors.

A2.2.5. Single model - Classification

It is possible for a model to have perfect calibration and good discrimination, but this may be less relevant if the discrimination benefits are driven by good performance at the extremes of the risk distribution, where this is unlikely to have any effects on clinical management decisions. This is because clinical management decisions are often made

at discrete risk thresholds, such as 1%, 5% and 10% absolute risk for CVD mortality. Therefore, the question of whether a participant's risk is 15% or 16% is merely academic and clinically irrelevant, as in both instances, this individual will be offered the same intervention. What is more important is how accurately models classify individuals into two or more predicted risk categories.

In the CVD example cited there may be as many as four categories stemming from three clinical thresholds (1%, 5%, and 10%). Usually one of these thresholds is perhaps the most important (e.g. the 5% thresholds). This is because the interventions offered to someone with a predicted risk of 4.9% is substantially different to the interventions offered to someone with a predicted risk of 5.1% (e.g. advised to start statins). Other instances, such as the development of cancer screening tests, also tend to use just one clinical threshold.

Many CVD clinical guidelines propose an additional complication: that statins are advised in case that predicted risk is high AND cholesterol readings are high. Although formal statistical techniques rarely discuss such two-by-two conditionality in the context of risk prediction, pragmatic articles in this area usually partition participants into two groups: those who meet and those who do not meet the wider set of conditions. This can be seen as an adaptation of the conventional classification process, with the caveat that the more conditionalities there are for participants to be designated as "for intervention", then the more this restricts the potential for superior risk prediction model to improve clinical reclassification.

When just one clinical threshold is deemed as important, then *Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV)* can be used to sufficiently describe classification performance. In simple terms, sensitivity means *"How many cases did it correctly predict?"*; specificity means *"How many controls did it correctly predict"*; PPV means *"If the test result reads 'high risk' for one individual, then what is the probability that they will become a case?"*; NPV means *"If the test*

result reads 'low risk' for one individual, then what is the probability that they will become a control?".

Sensitivity and Specificity are more useful at the perspective of the population (e.g. to those optimizing population-level resource allocation), while PPV and NPV are more useful at the perspective of the individual undergoing the test, who may ask their clinician *"Now that you have told me how my risk is 'high', what is the probability that you are correct – i.e. that I will indeed develop the disease?"*

A2.2.6. Single model – Classification thresholds and uptake

In most clinical instances, the PPV is lower than 50% and often lower than 10%. For example, in the context of predicting 10-year CVD risk in asymptomatic individuals, designation of "high risk", where absolute risk is greater than 5%, means that most people have an absolute risk between 5% to 10%. In other words, the clinician will have to typically respond to the patient that *"There is a 90% to 95% probability that I am wrong, and that you will not develop CVD in the next 10 years"*. This level of precision is poor, indicating how individual risk prediction science is still in its infancy, with much more progress to be made than has been made so far. However, a 10-year risk of 6% may still translate into a lifetime risk of 60%. Secondly, these interventional thresholds were initially about four times more stringent (e.g. 20% instead of 5%, when ASSIGN was developed in 2007(186)). As the cost of preventative interventions like statins has fallen, as well as emerging evidence of a better benefits-to-side effect profile than initially thought,(256) this has led to a decline in the absolute threshold at which clinical interventions are deemed to be cost-effective and recommended. Notably however, the decision of whether a patient ought to begin statin therapy once their absolute predicted risk exceeds 5% or 10% is heavily dependent on the disutility that a patient attributes to the burden of having to take a daily pill. This may well show large inter-individual variation. As a result, patients are increasingly being encouraged to be told not just that their risk is "high" but the actual absolute risk (i.e. "6%"), and to assist them in weighing the pros and cons

of statin therapy with information like "*for other people like yourself [who are above a predefined criteria of 5%], studies have found the benefits of statins to outweigh the costs and side effects, however it depends much on how much you mind having to take a daily pill".*

In some respects, this discussion, of moving away from dichotomous classifiers back towards continua of risk gradients has brought this introductory discussion full circle. Continua of risk are the mathematical basis of all statistical risk prediction models. These should be converted into dichotomous classifications, particularly to advise on cost-effectiveness and to form guidance around average practice for a large group of patients with similarly predicted risk. Thereafter, however, the final decision to proceed or decline personalized interventions is nonetheless personal, subjective and holistic, beyond the realm of statistical modelling. This withstanding, if the uptake of a given intervention in a population is substantially lower than what is predicted as optimal from statistical models, then this suggests either than clinicians should intervene more aggressively, or that the statistical models and guidance be updated to better reflect missing real-life disutilities of intervention.

A2.2.7. Single model – Classification trade-offs, and Net Benefit

The present discussion of classification has posited that there are two fundamental axes which are irreconcilable (e.g. sensitivity-axis Vs. specificity-axis if viewed from the perspective of the population; or PPV-axis Vs. NPV-axis if viewed from the perspective of the individual). Under the assumption of infinite clinical thresholds, discrimination statistics such as C-statistic posit that a one-unit increase in sensitivity is equivalent to a one-unit increase in specificity. In other words, that the losses of failing to identify one True Positive (who will hence forego a potentially useful intervention) is equivalent to the losses of identifying an additional False Positive (who will be subject to increased harm from needless intervention). In reality, the benefit-to-harms ratio of most interventions is rarely close to one. This is apparent in how most clinical interventional

thresholds are not set to 50% of the predicted individual probability, but are much lower than this to reflect the large potential benefits and small potential harms.

This can be illustrated with the hypothetical example of three healthy participants, each with a 10-year CVD mortality prediction of 1%, 5%, and 10% respectively. Virtually all clinicians and public health practitioners will agree, that no participants at 1% risk will benefit from statin therapy. Virtually all clinicians and public health practitioners will agree, that all participants at 10% risk will benefit from statin therapy. That is, they will willingly treat 10 such high-risk individuals for the benefit of 1 participant, despite the fact that 9 out of 10 of these participants will seemingly be treated needlessly. From these two extremes, we can infer that the cost-to-benefit ratio of statin use is smaller than 10:1 (latter example), and larger than 100:1 (former example).

If clinicians are presented with a hypothetical cohort of participants with 5% risk, then this may be the point of greatest ambivalence. That is, half the clinicians may err towards treating while half will err towards not treating. Having identified this, we can infer that the cost-to-benefit ratio of statins is approximately 20:1. In other words, 20 units of needless cost in 20 newly over treated participants (i.e. the combined total of side effects, financial cost, as well a burden of taking a daily pill and being labelled as unhealthy) are equal to 1 unit of benefit in one participant who has been newly identified correctly as being at risk. This observation can be exploited to evaluate risk prediction models, by inferring that the benefit gained from correctly identifying one additional True Positives is approximately 20x as important, when compared to the cost of incorrectly labelling one additional person as a False Positive. This was first noted by Peirce in an article published in *Science* in 1884, to allow for the calculation of a weighted *Net Benefit* of using a risk prediction model.(287) The formulation was extended by Vickers in 2006, who developed a graphical method of evaluating Net Benefit across a range of cost-to-benefit ratios,(255) which in my CVD example might stretch from 100:1 to

10:1. The underlying rationale is that when the true cost-to-benefit ratio is 100:1, a strategy of *"Treat nobody. There is no need to predict risk"* is preferred, as this results in greater Net Benefit. When the true cost-to-benefit ratio is 10:1, a strategy of *"Treat everybody. There is no need to predict risk"* is preferred, as this results in greater Net Benefit. Graphical methods, called *Decision Curve Analysis*, can determine the range of cost-to-benefit ratios, for which risk prediction is deemed superior to these two extreme examples where no risk prediction takes place. For example, if Decision Curve Analysis suggest that the Net Benefit of the risk prediction approach is superior when then cost-to-benefit ratio is smaller than 80:1 but larger than 12:1, then this can be transformed to ask *"is it plausible, that the risk thresholds at which clinicians are maximally ambivalent about intervention is somewhere between 1.3% and 8%?"* If the answer is yes, then this denotes that use of the risk prediction strategy may be warranted. Ideally, this would be followed by formal modelling of cost-effectiveness, which in turn would benefit from data taken from randomized trials. I have also not seen the use of metrics such as the *Number Needed to Treat* and/or the *Number Needed to Screen*. Both are well established in other applied domains of epidemiology (e.g. drug development and cancer screening, respectively), but have been seldom used in the literature on cardiovascular risk prediction. These could potentially assist in translating the utility of a model to a wider range of readers, who may not otherwise be able to discern whether the wider programme appears beneficial enough to implement, or not.

A2.2.8. Comparing two models – what to change in model 2

In CVD risk prediction settings, the baseline model that researchers typically try to improve already contains a number of predictors, some of which like serum cholesterol, are both invasive and time-consuming to collect. A little-researched area of risk prediction might be the demonstration of noninferiority: that the substitution of an invasive and time-consuming test in $model_1$, for a non-invasive and instant test in $model_2$ might show similar discrimination performance, at least in a pre-screen first stage approach.

By extension, most prediction models and screening tests are applied in a two-or-three step design. The first step is stratification by age and sex alone, leading to invitation. The second step is typically the face-to-face collection of less invasive data (incl. cholesterol for CVD, mammography for Breast Cancer, and PAP test for Cervical Cancer). The third step is either intervention (e.g. statins), more invasive testing (e.g. biopsy), or more frequent testing (shorter recall time between subsequent PAP tests). Most research efforts have concentrated on improving performance in the second and third stages. However, many countries hold substantial patient-level data that could also improve the automatic selection that takes place before invitation. For example, postcode could be additionally considered alongside age and sex, so that people living in more deprived areas with greater risk are invited for screening at an earlier age, when compared with people living in more affluent areas. Similar adjustments could potentially be made after considering existing data on marital and employment status, as well as data from electronic healthcare records for known comorbidities.

A2.2.9. Comparing two models – calibration, discrimination, and net benefit

I am not aware of a single analytical technique which directly compares the *calibration* performance of two competing models. Instead, calibration is best assessed visually.

For comparing *discrimination*, one common approach is to subtract the two C-statistics from one another, to derive ΔC. Some authors also offer a confidence interval around ΔC, but Harrell himself finds this inappropriate. Of note, if the baseline $model_1$ being improved is already a very poor model (e.g. with a C-statistic of 0.60), then it is much easier for an updated $model_2$ to improve this to say 0.65, when compared to the more difficult challenge of updating a good $model_1$ with C-statistic of 0.80 to an updated $model_2$ with a C-statistic of 0.85. This phenomenon can be thought of as “low hanging fruit” – the first predictors substantially improve

discrimination, while demonstrating additional benefit from “higher hanging fruit” is statistically more challenging.

To assess improvements in classification, one could inspect Δsensitivity and Δspecificity (if taking the population view) or ΔPPV and ΔNPV (if taking the individual view). As explained section A2.2.7, these two axes can be synthesized, by applying a cost-to-benefit ratio, to derive a *Net Benefit* statistic. This downweighs potential costs, arising from more False Positives and loss of specificity and PPV, while upweighing potential benefits, arising from more True Positives and improvements to sensitivity and NPV.

The *ΔNet Benefit* metric is perhaps one of the best all-rounder metrics to evaluate whether a new prediction tool adds incremental clinical value or not. However, its shortcoming is the fact that different clinical questions have different scales of Net Benefit, which is capped by the disease prevalence. Hence a Net Benefit of 0.05 might be seen as the peak for CVD models, while for rare diseases the peak could be as low as 0.0001. Furthermore, since ΔNet Benefit does not have a confidence interval, it is difficult to judge whether a ΔNet Benefit of 0.001 is clinically meaningful or not, and whether practitioners ought to change practice. Ideally, a formal cost-effectiveness analysis could model whether spending more funds on a new risk prediction approach is financially warranted, and these could be explored in all instances where ΔNet Benefit exceeds zero.

A range of ΔNet Benefit scenarios can be investigated, by varying the cost-to-benefit weighting in a Decision Curve Analysis (figure A2-1). Despite the strong theoretical basis for *Net Benefit* and *Decision Curve Analysis*, these rarely been applied to risk prediction papers to date.

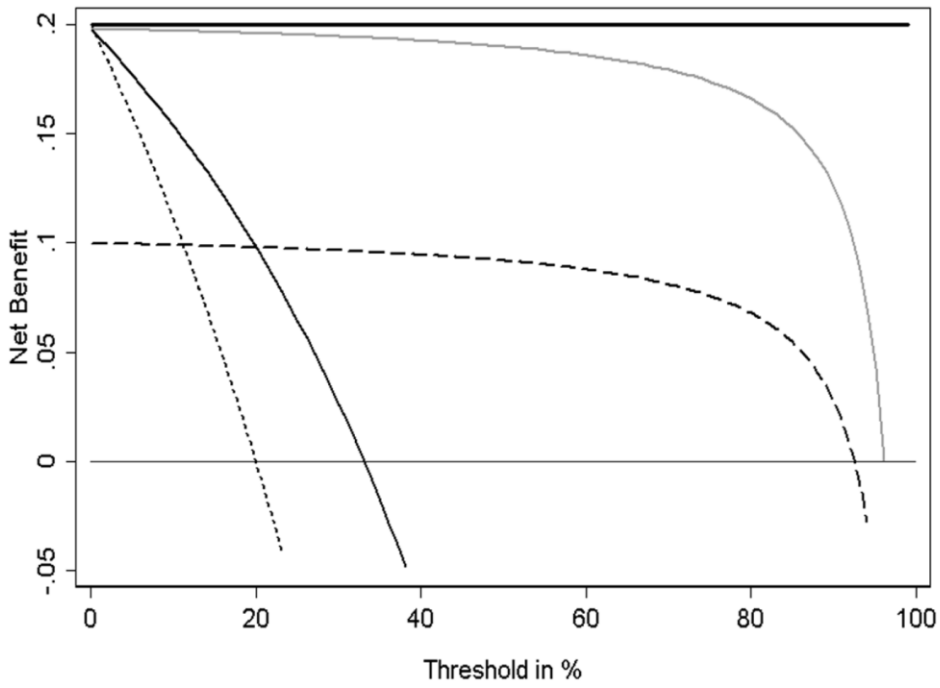


Figure A2-1. Theoretical examples of decision curve analyses.
Thin line: assume no patient has disease (treat none)
Dotted line: assume all patients have disease (treat all)
Thick line: a perfect prediction model.
Grey line: a near perfect binary predictor (99% sensitivity and 99% specificity). Solid line: a sensitive binary predictor (99% sensitivity and 50% specificity). Dashed line: a specific binary predictor (50% sensitivity and 99% specificity).

If clinical guidelines advise interventions at some threshold which is between 0 to 25%, then sensitive predictors and models are preferred over specific predictors for that application. Adapted from (255).

A2.2.10. Comparing two models – net reclassification and its critique

A popular metric which is used to compare two risk prediction models is the Net Reclassification Index (NRI) and its various formulations. This is the proportion of cases now classified better with the new model, plus the proportion of controls now classified better with the new model. The original and most common formulation assumes an infinite number of clinical risk thresholds, also known as *Continuous NRI*.

Theoretically, as NRI is the sum of two proportions (each with a range of 0-100%), the maximum NRI value is 2. Summary NRI values themselves should never be presented or interpreted as proportions or percentages, although a recent review showed how most published examples fail to follow this advice.(288)

Despite the popularity of *Continuous NRI* in published risk prediction papers, the construct contains considerable limitations, as detailed by Kerr(289):

1) It is not clear why one would want to allocate 50% **weighting** to improvement seen among cases, and another 50% weighting to improvement seen among controls. A population-based approach might instead weigh the case improvement by the prevalence of disease (e.g. 5%), and the control improvement by the prevalence of controls (e.g. 95%). Perhaps the allocation of more weight can be seen to reflect the small cost-to-benefit ratio of most interventions. However, Continuous NRI is not formally weighted as such, so the 50:50 allocation can rightly be criticized as arbitrary.
2) *Continuous NRI* is larger, if the baseline model being improved already has better **discrimination**. For example, a continuous NRI of 0.62 is just as compatible with a small C-statistic improvement from 0.900 to 0.919, as it is with a large C-statistic improvement from 0.700 to 0.780. In other words, imprecise models have to work harder to demonstrate improvements in Continuous NRI. This is a function of how precise models already discriminate well between cases and controls, and so slight amplification of this risk divergence between these two groups leads to large apparent reclassification.
3) *Continuous NRI* is larger, if the baseline model being improved is poorly **calibrated** to the validation dataset, where improvements are tested.
4) The unit of measure is binary movement, either up or down in predicted risk, without consideration of the **magnitude** of this change. For example, if the predicted risk of a case moves from 6.0% to 6.1%,

then this is equivalent to if the predicted risk of a case moves from 6.0% to 60.0%.

Many authors have recommended to refrain from presenting the combined *Continuous NRI*. If the clinical example being studied truly contains no categories of risk thresholds for intervention, then it may be more appropriate to present the percentage reclassification among cases as one statistic, and the percentage reclassification among controls as another statistic. However, in the case of CVD prediction, Categorical formulations are preferred.

Categorical NRI creates more stringent conditions before a change in risk prediction between two models "counts". This makes their values smaller than the values of Continuous NRI. In cases where there are at least three categories (i.e. at least two thresholds):

5) *Categorical NRI* can easily be increased by increasing the number of risk thresholds.
6) Classifying cases as "high → intermediate risk" is given equivalent penalty, as is classifying cases from "high → low risk". Treatment decisions for these two cases could diverge, which is not captured in the indiscriminate way that intermediate changes are handled. Reclassification tables have been proposed as a more granular way to presenting such data.

Finally, one can consider a special case of Categorical NRI, which contains just two categories (i.e. one threshold). I will call this *Binary NRI*. The main limitation here is the same as number 1 above – that improvements to cases and controls are weighted equally (agnostic of the true prevalence of disease).

An adaptation of Binary NRI is to up-weight the improvement in cases, and down-weight the improvement in controls. These formulations do not usually adjust for the prevalence of case and control status. This may have the effect of giving excess attention to improvements in cases,

without giving due attention to potential deterioration among controls. i.e. it risks biasing the final weighted-binary-NRI measure towards models that are very sensitive, but lack specificity, thereby propagating overtreatment.

By formally accounting for both the cots-to-benefit ratio of undertreating cases and overtreating controls, as well as the prevalence of the case: control mix, we have come full circle back to a more simplified metric: the *ΔNet Benefit*. Therefore, in clinical decisions where there is often one main risk threshold (such as recommending statins for primary CVD prevention), one might be best placed to avoid the NRI construct altogether. On the other hand, NRI statistics remain popular with editors and peer-reviewers, so a cautious approach would be to include these as a sensitivity analysis.

Annex 3 – Peer-reviewed papers

A3.1. Mendelian randomization

Education and coronary heart disease: mendelian randomisation study

Taavi Tillmann,[1] Julien Vaucher,[2] Aysu Okbay,[3] Hynek Pikhart,[1] Anne Peasey,[1] Ruzena Kubinova,[4] Andrzej Pajak,[5] Abdonas Tamosiunas,[6] Sofia Malyutina,[7,8] Fernando Pires Hartwig,[9,10] Krista Fischer,[11] Giovanni Veronesi,[12] Tom Palmer,[13] Jack Bowden,[10] George Davey Smith,[10,14] Martin Bobak,[1] Michael V Holmes,[10,15,16,17]

For numbered affiliations see end of article.

Correspondence to: T Tillmann t.tillmann@ucl.ac.uk

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2017;358:j3542 http://dx.doi.org/10.1136/bmj.j3542

Accepted: 09 July 2017

ABSTRACT

OBJECTIVE

To determine whether educational attainment is a causal risk factor in the development of coronary heart disease.

DESIGN

Mendelian randomisation study, using genetic data as proxies for education to minimise confounding.

SETTING

The main analysis used genetic data from two large consortia (CARDIoGRAMplusC4D and SSGAC), comprising 112 studies from predominantly high income countries. Findings from mendelian randomisation analyses were then compared against results from traditional observational studies (164 170 participants). Finally, genetic data from six additional consortia were analysed to investigate whether longer education can causally alter the common cardiovascular risk factors.

PARTICIPANTS

The main analysis was of 543 733 men and women (from CARDIoGRAMplusC4D and SSGAC), predominantly of European origin.

EXPOSURE

A one standard deviation increase in the genetic predisposition towards higher education (3.6 years of additional schooling), measured by 162 genetic variants that have been previously associated with education.

MAIN OUTCOME MEASURE

Combined fatal and non-fatal coronary heart disease (63 746 events in CARDIoGRAMplusC4D).

RESULTS

Genetic predisposition towards 3.6 years of additional education was associated with a one third lower risk of coronary heart disease (odds ratio 0.67, 95% confidence interval 0.59 to 0.77; $P=3\times10^{-8}$). This was comparable to findings from traditional observational studies (prevalence odds ratio 0.73, 0.68 to 0.78; incidence odds ratio 0.80, 0.76 to 0.83). Sensitivity analyses were consistent with a causal interpretation in which major bias from genetic pleiotropy was unlikely, although this remains an untestable possibility. Genetic predisposition towards longer education was additionally associated with less smoking, lower body mass index, and a favourable blood lipid profile.

CONCLUSIONS

This mendelian randomisation study found support for the hypothesis that low education is a causal risk factor in the development of coronary heart disease. Potential mechanisms could include smoking, body mass index, and blood lipids. In conjunction with the results from studies with other designs, these findings suggest that increasing education may result in substantial health benefits.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Many observational studies have found that people who spend more time in educational settings subsequently develop less coronary heart disease

However, whether this association is causal is not clear, partly because randomised controlled trials are practically infeasible in this area

Few studies have applied mendelian randomisation to investigate how exposure to socioeconomic risk factors might causally change the risk of disease occurrence

No such study has done sensitivity analyses around genetic pleiotropy

WHAT THIS STUDY ADDS

Increasing the number of years that people spend in the educational system may lower their risk of subsequently developing coronary heart disease by a substantial degree

These findings should stimulate policy discussions about increasing educational attainment in the general population to improve population health

Introduction

Coronary heart disease (CHD) is the leading cause of death globally. Whereas the causal effects of risk factors such as smoking, high blood pressure, and raised low density lipoprotein cholesterol are generally accepted and reflected in disease prevention strategies, substantial uncertainty still surrounds other potential factors. Decades of observational studies have consistently associated socioeconomic factors such as higher education with decreased risk of CHD.[1-4] However, this association may not stem from an underlying causal effect but may arise owing to the methodological limitations of traditional observational research.[5 6] Clarifying whether the association between education and CHD is causal has widespread implications for our understanding of the causes of CHD, as well as for the potential development of novel population based approaches to its prevention. Unfortunately, randomised controlled trials are practically infeasible in this area, given the long (approximately 50 year) interval between exposure and outcome. Improving causal inference through other study designs is therefore necessary.

Mendelian randomisation analysis uses genetic variants associated with a risk factor (for example,

education) to make causal inferences about how environmental changes to the same risk factor would alter the risk of disease (for example, CHD).[7] Comparing the risk of disease across participants who have been grouped by their genotype enables the causal effect of a risk factor to be approximated with substantially less bias than in a traditional observational analysis. Genetic markers of a risk factor are largely independent of confounders that may otherwise cause bias, as genetic variants are randomly allocated before birth.[8] This, as well as the non-modifiable nature of genetic variants, provides an analogy to trials, in which exposure is allocated randomly and is non-modifiable by subsequent disease.[8]

Until relatively recently, mendelian randomisation analyses have been conducted on single datasets in which data on genotype, risk factor, and outcome were measured for all participants (known as "one sample mendelian randomisation"). However, advanced analyses on pleiotropy require larger sample sizes to maintain statistical power. This would require data pooling across dozens of studies, which is administratively difficult to organise. As an alternative, summary level data from large genome-wide associations study (GWAS) consortia have become increasingly available in the public domain. Such data can be used to conduct mendelian randomisation analyses, whereby gene exposure measures are taken from one GWAS and gene outcome measures are taken from another GWAS (altogether known as "two sample mendelian randomisation").[9] Further methodological developments, including mendelian randomisation-Egger (commonly abbreviated to MR-Egger), weighted median mendelian randomisation, and mode based methods, can all be used as sensitivity analyses to additionally investigate any pleiotropic effects of the genetic variants (that is, when genetic variants for education exert their influence on heart disease through an "off-target" pathway that bypasses the education phenotype; see supplementary figure 1 for details.[9101011] The mendelian randomisation method has successfully been applied to a range of biological and behavioural exposures.[12 13] We are aware of just two studies that have applied it to investigate a socioeconomic exposure: a polygenic score for education has previously been associated with the development of myopia and dementia.[14 15] However, these studies did not investigate the possibility of genetic pleiotropy.

Our primary research question was "Is there genetic support for the hypothesis that education is a causal risk factor in the development of CHD, and, if so, does education cause changes to conventional cardiovascular risk factors that could be mediators of this?" We firstly updated traditional observational estimates of the association between education and risk of CHD from several large studies and consortia. Secondly, we applied two sample mendelian randomisation analyses to investigate whether people with a genetic predisposition towards higher education have a lower risk of CHD. A recent GWAS from the Social Science Genetic Association Consortium (SSGAC) identified a large number of independent genetic variants (single nucleotide polymorphisms—SNPs) associated with educational attainment.[16] We used 162 such SNPs to mimic the process of randomly allocating some participants to more education and other participants to less education. To compare the CHD risk of participants randomised in such a manner, we then used data from the Coronary Artery Disease Genome wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics Consortium (CARDIoGRAMplusC4D) to see whether participants with genetic variants for longer education had an altered risk of CHD compared with participants with genetic variants for shorter education.[17] Careful consideration of the results from such analyses, as well as the wider literature, can support inferences about the likely cardiac consequences from environmentally acquired alterations to education. We checked the robustness of our findings across a range of sensitivity analyses and additionally tested for reverse causation by checking whether those SNPs that best predict CHD also associate with educational outcomes. Supplementary figure 2 illustrates the mains steps taken in this study.

Methods

Throughout all analyses, we defined education in the same way as in the original GWAS analysis, in which data from 65 studies were harmonised against the International Standard Classification of Education 1997 classification system (see supplementary table 1.3 of the original GWAS study[16]). After harmonisation, self reported educational attainment was modelled linearly, expressed as one standard deviation (that is, 3.6 years) of additional schooling. In this form, one year of vocational education was equivalent to one year of academic education, and we did not assume any qualitative differences in the type of education. We defined CHD as a composite of myocardial infarction, acute coronary syndrome, chronic stable angina or coronary stenosis of more than 50%, or coronary death.

Observational association between education and CHD

In traditional observational analysis, we used a combination of cross sectional and prospective data, collected between 1983 and 2014 (table 1). For prevalent CHD cases in cross sectional data, we analysed 43 611 participants (1933 cases) from the National Health and Nutrition Examination Surveys (NHANES) (see supplementary figure 3).[26] For incident CHD cases in prospective data, we analysed 23 511 participants (632 cases) from the Health, Alcohol and Psychosocial factors In Eastern Europe (HAPIEE) study[18] and combined this with published estimates from 97 048 participants (6522 cases) of the Monica Risk, Genetics, Archiving and Monograph (MORGAM) study in Europe (see supplementary table 1 for case definitions and statistical details).[3 19]

doi: 10.1136/bmj.j3542 | BMJ 2017;358:j3542 | the bmj

Table 1 | Details of studies and datasets included in analyses

Analysis/study	Risk factor/outcome	Participants (CHD cases)	Web source (if publicly available)
Traditional observational analysis			
NHANES	Years of education/non-fatal CHD	43 611 (1933)	www.cdc.gov/nchs/nhanes/
HAPIEE[18]	Years of education/fatal and non-fatal CHD	23 511 (632)	–
MORGAM[19]	Years of education/fatal and non-fatal CHD	97 048 (6522)	–
Mendelian randomisation analysis (education to CHD and CHD to education)			
SSGAC[16]	Years of education	349 306	www.thessgac.org/data
CARDIoGRAMplusC4D[17]	CHD	194 427 (63 746)	www.cardiogramplusc4d.org/data-downloads/
Mendelian randomisation analysis (education to conventional cardiovascular risk factors)			
TAGC[20]	Smoking	74 053	www.med.unc.edu/pgc/results-and-downloads
ICBP[21]	Blood pressure	74 064	www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000585.v1.p1
GLGC[22]	LDL cholesterol, HDL cholesterol, and triglycerides	188 577	csg.sph.umich.edu/abecasis/public/lipids2013/
DIAGRAM[23]	Type 2 diabetes	149 821	diagram-consortium.org
MAGIC[24]	Glucose	133 010	www.magicinvestigators.org
GIANT[25]	Body mass index, height	339 224	portals.broadinstitute.org/collaboration/giant/

CARDIoGRAMplusC4D=Coronary Artery Disease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus the Coronary Artery Disease (C4D) Genetics consortium; CHD=coronary heart disease; DIAGRAM=Diabetes Genetics Replication and Metaanalysis; GIANT=Genetic Investigation of Anthropometric Traits; GLGC=Global Lipids Genetic Consortium; HAPIEE=Health, Alcohol and Psychosocial factors in Eastern Europe; HDL=high density lipoprotein; ICBP=International Consortium for Blood Pressure; MAGIC=Meta-Analyses of Glucose and Insulin-related traits Consortium; LDL=low density lipoprotein; MORGAM=Monica Risk, Genetics, Archiving and Monograph; NHANES=National Health and Nutrition Examination Survey; SSGAC=Social Science Genetic Association Consortium; TAGC=Tobacco and Genetics Consortium.

Genetic variants associated with education

We retrieved a shortlist of SNPs associated with educational attainment from a recent GWAS involving 405 072 people of European ancestry (table 1).[16] For our main analysis, we used 162 independent SNPs associated ($P<5.10^{-8}$; linkage disequilibrium $r^2<0.1$) with education in a meta-analysis of the discovery (SSGAC) and replication (UK Biobank) datasets. Altogether, these 162 SNPs explained 1.8% of the variance in education. This is sufficient to generate a strong genetic instrument with which to derive unbiased causal estimates (see supplementary table 2 for power calculations). For our secondary analysis, we used another set of 72 independent SNPs (at $r^2<0.1$) that were associated with education in the discovery dataset (SSGAC) alone (293 723 participants; $P<5.10^{-8}$) and that were subsequently found to be directionally consistent in an independent replication dataset (UK Biobank; see supplementary figure 4 for a summary of how SNPs were selected). We decided to use the larger set of instruments (with 162 SNPs) in our main analysis instead of the smaller set of instruments (with 72 SNPs) to maintain sufficient statistical power for our sensitivity analyses. To avoid potential biases that may arise when datasets contributing towards the SNP-to-exposure and SNP-to-outcome estimates overlap, we excluded studies in SSGAC that overlapped with CARDIoGRAMplusC4D (full details of these excluded studies are provided in supplementary methods 3.1). We then checked that the removal of these overlapping datasets from SSGAC had no material effect on the SNP-to-education estimates (see supplementary figures 5 and 6 for further details).

Genetic variants associated with CHD

Data on CHD have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.cardiogramplusc4d.org. For each of the 162 SNPs associated with education, we retrieved summary level data for either the same SNP (115 of 162 SNPs) or for a proxy SNP in high linkage disequilibrium (47 of 162 SNPs at $r^2>0.8$) from datasets totalling 63 746 CHD cases and 130 681 controls (see supplementary figure 7 for how the education SNPs were matched against the CHD GWAS dataset).[17] We repeated a similar process for our secondary analysis using a set of 72 SNPs (supplementary figure 8).

Statistical analyses

Traditional observational analyses

We used Cox proportional hazards and logistic regressions to calculate traditional observational estimates for incident and prevalent cases, respectively. Results were adjusted for age and sex. Further methodological details are given in supplementary methods 1.

Mendelian randomisation analyses

For all mendelian randomisation analyses, alleles from the SSGAC and CARDIoGRAMplusC4D datasets were aligned to correspond to an increase in educational attainment. To investigate whether education is likely to play a causal role in coronary heart disease, we used three mendelian randomisation approaches. Firstly, we used conventional (also termed 'inverse variance weighted") mendelian randomisation analyses, by regressing the SNP-education associations (exposure) against the SNP-CHD associations (outcome), with each SNP as one data point (details in supplementary methods 3.1).

Secondly, we used three sensitivity analyses to investigate to what degree pleiotropic effects might bias the mendelian randomisation causal estimates. These methods allow some of the mendelian randomisation assumptions to be relaxed. For example, mendelian randomisation-Egger relies on the InSIDE assumption, which requires that the magnitude of any pleiotropic effects (from SNPs to CHD, which bypasses education) should not be correlated with the magnitude of the main effect (from SNP to education).[10] Median based and mode based methods posit that when looking at lots of SNPs (some of which may have pleiotropic

effects on CHD), these pleiotropic effects are likely to be comparatively heterogeneous in nature and hence less likely to converge on a common median/modal estimate. In contrast, valid SNPs with no pleiotropic effects are more likely to show more uniform and homogeneous effects (on education and thereafter CHD), which makes them more likely to cluster towards the median/modal point estimate.[9 27] These methods are fully described in supplementary methods 3.2. Consistency of results across a range of methods that make different assumptions about pleiotropy strengthens causal inference, whereas divergent results may indicate that genetic pleiotropy is biasing some of these results (described in supplementary figure 1).

Thirdly, to check whether genetic risk for coronary events might be a causal factor for educational attainment, we did mendelian randomisation in the opposite direction (bidirectional mendelian randomisation) using 53 SNPs associated with CHD (supplementary methods 3.2.4). Under conditions of massive pleiotropy, genetic risk of coronary events might also predict educational outcomes.

To investigate potential mechanisms from education to CHD, we applied conventional mendelian randomisation to investigate whether genetic predisposition towards longer education could lead to improvements in the established cardiovascular risk factors. In this analysis, we discarded 60 SNPs with missing data on one of the cardiovascular risk factors from the 162 SNP instrument and thus used a smaller set of 102 SNPs (details in supplementary methods 3.3 and supplementary figure 4).

Patient involvement

Patients were not involved in the design or implementation of this study. There are no specific plans to disseminate the research findings to participants, but findings will be returned back to the original consortia, so that they can consider further dissemination.

Results

Observational analyses

On the basis of NHANES data, each additional 3.6 years of education (1 SD) was associated with 27% lower odds of prevalent CHD (odds ratio 0.73, 95% confidence interval 0.68 to 0.78; illustrated in figure 1). In prospective analyses, 3.6 years of additional education was associated with a 20% lower risk of incident CHD in the HAPIEE and MORGAM studies, with a pooled hazard ratio of 0.80 (0.76 to 0.83). Cohort specific results from MORGAM are additionally shown in supplementary figure 9.[18 19] These observational estimates were robust to sensitivity analyses accounting for different case definitions, age at first CHD event, and potential confounding by other measures of socioeconomic position (supplementary table 3). We also saw evidence for a dose-response relation between the amount of education and risk of CHD (supplementary figures 10 and 11).

Genetic association between education and CHD

After integrating two GWAS datasets and examining millions of SNPs across the entire genome, we found strong evidence for a negative genetic correlation between education and CHD (r_g=−0.324; r_g^2=0.104; P=2.1×10^{-12}; further details in supplementary methods 2).[28] To interpret this, educational outcomes can vary as a result of genetic and non-genetic variance. Within the domain of genetic variance, approximately 10% of the genetic variance of education seems to be shared with the genetic variance of CHD, whereby this correlation is negative. This correlation can arise for various reasons, so we next did multiple mendelian randomisation analyses to investigate the presence and direction of any causal effects.

Causal effect from education to CHD

Using conventional mendelian randomisation analysis, 1 SD longer education (due to genetic predisposition across 162 SNPs) was associated with a 33% lower risk of CHD (odds ratio 0.67, 0.59 to 0.77; P=3×10^{-8}). Supplementary figure 12 additionally shows individual causal estimates from each of the 162 SNPs. As expected, sensitivity analyses using mendelian randomisation-Egger and weighted median mendelian randomisation provided less precise estimates than with conventional mendelian randomisation. Nonetheless, their causal estimates were similar in terms of direction and magnitude, and they were unlikely to have happened by chance alone (fig 1). We found little evidence of a non-zero intercept from the mendelian randomisation-Egger test (intercept β=0.004, −0.056 to 0.013; P=0.417), consistent with the hypothesis that genetic pleiotropy was not driving the result. The mendelian randomisation regression slopes are illustrated in supplementary figures 13 and 14. A secondary set of analyses using a set of 72 SNPs instead of 162 SNPs yielded consistent results in terms of direction and magnitude (fig 1).

Further sensitivity analyses, using both sets of instruments, are reported in supplementary table 4. Briefly, an analysis that can account for some measurement error in our genetic instruments for exposure (so-called mendelian randomisation-Egger+SIMEX) gave similar findings.[29] Results from modal based mendelian randomisation approaches were consistent with the hypothesis that genetic pleiotropy was not driving the conventional mendelian randomisation result. We also did robustness checks by omitting SNPs with higher levels of missing data, as well as SNPs that were available in the CHD GWAS dataset in the form of a proxy SNP. These gave similar results in terms of direction, magnitude, and statistical significance. Collectively, all these sensitivity analyses make it less likely that the presence of pleiotropic effects, or missing data, grossly biased our main causal analysis.

Causal effect from CHD to education

We found little evidence for the hypothesis that genetic liability for CHD risk is associated with educational outcomes. Namely, 1-log greater genetic risk of CHD

doi: 10.1136/bmj.j3542 | BMJ 2017;358:j3542 | thebmj

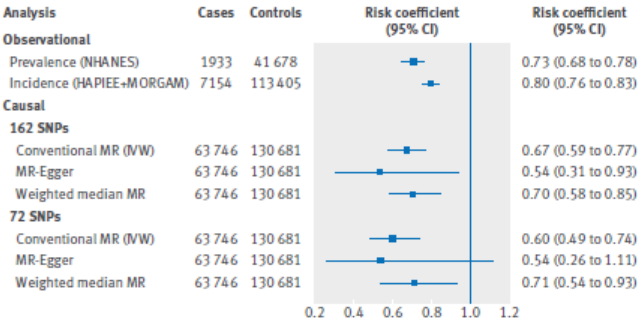


Fig 1 | Comparison of observational and causal estimates for risk of coronary heart disease (CHD), per 3.6 years of educational attainment. Two observational estimates are provided according to prevalent and incident CHD cases. Risk coefficient for observational incident cases was derived by meta-analysis of hazard ratios from Health, Alcohol and Psychosocial factors In Eastern Europe (HAPIEE) and Monica Risk, Genetics, Archiving and Monograph (MORGAM) studies. Risk coefficients for observational prevalent cases and six causal estimates from mendelian randomisation (MR) are all odds ratios (see supplementary methods for full description of each analysis). IVW=inverse variance weighted approach; NHANES=National Health and Nutrition Examination Survey

was associated with 2.4 (−16.6 to 21.4) days of longer educational attainment. Results were unchanged after application of mendelian randomisation-Egger and weighted median mendelian randomisation (fig 2). The results from individual SNPs are shown in supplementary figures 18-20.

Causal effect from education to cardiovascular risk factors

To identify potential risk factors that could mediate the association between education and CHD, we investigated whether genetic predisposition towards longer education was associated with established cardiovascular risk factors. Table 2 shows that, in conventional mendelian randomisation analyses, a 1 SD longer education (due to genetic predisposition across 102 SNPs) was associated with a 35% lower odds of smoking, 0.17 lower body mass index, 0.14 mmol/L lower triglycerides, and 0.15 mmol/L higher high density lipoprotein cholesterol, with a P value smaller than 0.001 for each of these four outcomes. Associations with diabetes and systolic blood pressure

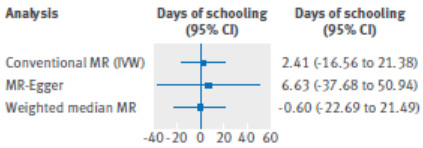


Fig 2 | Association of genetic liability to coronary heart disease (CHD) (exposure) on numbers of days of schooling (outcome). Causal estimates are expressed as difference in days of education per 1-log unit increase in risk of CHD as instrumented by 53 SNPs. Supplementary methods 3.2 details each mendelian randomisation (MR) analysis. IVW=inverse variance weighted approach

were in the anticipated direction, but these effects may have been due to chance or insufficient statistical power (P values 0.05 to 0.08).

Discussion

In this mendelian randomisation study, we found strong genetic support for the hypothesis that longer education has a causal effect on lowering the risk of coronary heart disease. Our findings using genetic data, which can be considered as "nature's randomised trials,"[30] were consistent with data from observational studies, and we found little evidence that our results may be driven by genetic pleiotropy. More specifically, 3.6 years of additional education (similar to an undergraduate university degree) is predicted to translate into about a one third reduction in the risk of CHD.

Comparison with previous studies

A vast body of observational studies across a range of settings show an association between education and CHD. In contrast, comparatively few studies have explicitly investigated the causality of this association. The existing studies on causality come from three domains. Firstly, analyses of natural experiments have compared mortality before and after changes to compulsory schooling laws—for example, by looking at mortality rates in countries before and after the introduction of national legislation that increased minimum education. In the Netherlands, such changes were associated with reductions in all cause mortality.[31] In the UK, the largest study so far reported causal effects on improving physical activity, body mass index, blood pressure, diabetes, CHD, and all cause mortality.[32] An extension of this design is to compare geographical areas, such as the various states in the US. These studies initially suggested a large effect on all cause mortality, but this effect disappeared when state specific baseline trends were taken into account.[33 34] In Sweden, an intervention to extend compulsory schooling throughout a 13 year transition period in a stepped wedge design across multiple municipalities reported lower all cause mortality in those deaths occurring after age 40 (equivalent to hazard ratio of death of 0.86 (0.77 to 0.96) per 3.6 years of additional education).[35]

Another source of causal inference comes from studies on monozygotic twins. Within each pair, both twins are exposed to the same set of genetic exposures (and also some environmental exposures, called the "shared environment"). Consequently, any difference in disease outcome between twins cannot arise from genetic effects. If differences in outcome associate with differential exposure to non-shared features of the environment (such as one twin pursuing education longer than the other twin), and if the magnitude of this association is comparable to that seen in the general population, this makes less likely the possibility that the observational association is confounded by genetic (or shared environmental) factors. Although the twin method does not eliminate the possibility of confounding from other factors in the

Table 2 | Causal effects from 3.6 years of education to 10 cardiovascular risk factors

Outcome	Causal effect (95% CI)*	P value
Binary traits		
Smoking status	0.65 (0.54 to 0.79)	≤0.001
Diabetes mellitus, type 2	0.75 (0.56 to 1.01)	0.057
Continuous traits		
Systolic blood pressure	−1.36 (−2.85 to 0.12) mm Hg	0.075
Diastolic blood pressure	−0.23 (−1.22 to 0.76) mm Hg	0.645
Low density lipoprotein cholesterol	−0.03 (−0.10 to 0.05) mmol/L	0.513
High density lipoprotein cholesterol	0.15 (0.07 to 0.23) mmol/L	≤0.001
Triglycerides	−0.14 (−0.22 to −0.06) mmol/L	≤0.001
Glucose	−0.02 (−0.08 to 0.03) mmol/L	0.441
Body mass index	−0.17 (−0.26 to −0.08)	≤0.001
Height	0.06 (−0.03 to 0.16) cm	0.208

All analyses are based on a common set of 102 single nucleotide polymorphisms associated with education, available in eight genome-wide association study consortia (see supplementary methods 3.3).
*Estimates are expressed as absolute values for continuous risk factors and as odds ratios for binary traits; both correspond to 1 standard deviation longer education (equivalent to 3.6 years of schooling).

non-shared environment, it is a design with which to eliminate the possibility of confounding from genetic factors. Twin studies conducted in Denmark initially found evidence both for and against causal effects from education to mortality and CHD incidence.[36 37] The largest study to date from Sweden (which has twice the statistical power of the previous largest study) found strong evidence for causal effects.[38] There, the association between years of education and lifespan did not attenuate at all when the conventional population based analysis was compared against the between twin analysis. Hence the twin literature suggests that, although only a handful of sufficiently powered studies exist, shared environmental factors (such as parenting) are not likely to cause substantial confounding. It also suggests that confounding from genetic factors (such as genetic differences in drive, motivation, personality, or innate intellect, all of which may predispose towards longer education) might not account for the observational associations between education and disease.

A parallel domain of research, using data from millions of non-identical siblings (that sometimes reached 100 times larger sample sizes than the twin studies), has also observed little attenuation of the association between education and subsequent mortality when comparing the general population analysis with the within sibling analysis.[39 40] As with twin studies, this also suggests that environmental and genetic factors shared by the siblings are unlikely to confound the observational association seen between education and disease. Although twin and sibling studies both leave open the possibility of confounding from non-shared environmental factors, taken together with our results (using an entirely different method), the wider body of evidence is more compatible with a causal interpretation, suggesting that increasing education may lead to a reduction in CHD.

Finally, some recent studies have also looked at specific genetic variants for education. An association was found between parental longevity and genetic markers for education in their offspring.[41] However, causal directions and pleiotropy were not tested in this study. Others have used conventional mendelian randomisation and found that genetic variants for education predict myopia and dementia.[14] However, these studies did not investigate pleiotropy of their genetic instruments. No mendelian randomisation studies of socioeconomic exposures have investigated any other disease outcome, such as cardiovascular diseases. Furthermore, most of the other designs listed above (including natural experiments and twin and sibling designs) have reported outcomes for all cause mortality. Few have reported cardiovascular mortality, and virtually none have reported fatal/non-fatal CHD, as we have.

Strengths and limitations

Our study has important strengths. We investigated the causality of the association between an easily measured socioeconomic factor (education) and a common disease (coronary heart disease). We applied the mendelian randomisation design, which in conjunction with findings from other study designs should improve our understanding of causality by reducing bias from confounding. By integrating summary level data from more than half a million individuals, our study was well powered to derive robust causal effect estimates and also powered for multiple sensitivity analyses (which typically require larger sample sizes). We used recent state of the art methodological developments to thoroughly explore the possibility of pleiotropy in our genetic variants, for which we found little evidence.

Our study also has some limitations. Firstly, the genetic variants associated with education may instead mark more generic biological pathways (such as vascular supply or mitochondrial function), which could enhance systemic fitness, thereby leading to parallel increases in cognitive and cardiac function.[42 43] Under this scenario, which violates the InSIDE assumption, policy interventions to increase education may not translate into lower incidence of heart disease. However, such a scenario is less likely to lead to the consistent set of results we found across our sensitivity analyses, as this would require that pleiotropy occurs in a scenario in which the InSIDE assumption is violated (so that mendelian

doi: 10.1136/bmj.j3542 | BMJ 2017;358:j3542 | thebmj

randomisation-Egger is biased), at least 50% of the information comes from SNPs with highly pleiotropic effects on heart disease, and these pleiotropic effects occurred in such a way as to make the causal estimates on heart disease seem very similar to one another. No definitive tests exist with which to verify such assumptions, meaning that triangulation of data from other sources and subjective judgment are needed to evaluate the plausibility of gross pleiotropic bias.[44] We believe such pleiotropy to be unlikely for four reasons. Firstly, the effects from genetic pleiotropy would have to coincide with the non-genetic associations observed in studies of monozygotic twins; secondly, they would also have to coincide with the non-genetic associations observed in natural experiments. Thirdly, if education and CHD share some of their underlying genome-wide genetic architecture (as seen in our LD score regression), and if most of the top hits for education are strongly pleiotropic for CHD, then one might imagine the top hits for CHD to also pick up some of these pleiotropic traits. However, our reverse direction mendelian randomisation found a null estimate. Fourthly, despite gaps in our understanding of the biological mechanisms through which these 162 SNPs influence education, they are disproportionately found in genomic regions that regulate brain development, they are enriched for biological pathways involved in neural development, and they are preferentially expressed in neural tissue.[16] As these 162 SNPs do not seem to have any expression or enrichment in cardiovascular tissues, this further narrows the scope for pleiotropy: any potential pleiotropy might have to exert a large effect on CHD via predominantly neurological pathways (for example, behaviours associated with obesity), rather than via global or systemic measures of fitness (such as mitochondrial function). Therefore, on balance, we believe that the scenario in which gross pleiotropy invalidates our sensitivity analysis is less consistent with the broader body of evidence, in comparison with the scenario in which our sensitivity analyses are valid. If our main and sensitivity analyses are valid, then policy interventions that mirror prolonged exposure to education (as indexed by our genetic instruments) should, on balance, probably prevent heart disease.

A second limitation is that to arrive at such a policy recommendation one would have to assume that genetic predisposition towards higher educational attainment causes the same behavioural and physiological consequences as environmentally acquired changes to educational attainment, such as from a policy intervention. It may be, however, that a year of additional education from genetic causes could trigger a different set of biological and behavioural mechanisms compared with a year of additional education resulting from policy change. We know very little about the mechanisms of these genetic effects. In the analyses we did in this study, we found some initial evidence that some of these genetic effects may be mediated via common cardiovascular factors such as smoking, body mass index, and lipids. In keeping with this, policy changes to education in the US and UK have also estimated some causal effects on smoking, body mass index, blood pressure, and diabetes,[32 45] which are broadly consistent with our findings. Few studies have measured the causal effects of policy interventions on blood lipids. Although a randomised controlled trial of education is difficult for CHD outcomes, owing to approximately 50 years of lag, future research using real life interventions may be able to measure effects on potential mediators, as these occur much sooner. A second response to this overall limitation is the analogy to other exposures (such as low density lipoprotein cholesterol and systolic blood pressure), for which genetic effects have mirrored findings from environmentally acquired changes (such as from randomised controlled trials of drug therapies.[4647]). Taken together, although our study makes no direct inference on what health effects may stem from a policy intervention that successfully increases education, we are cautiously optimistic that such a policy should lead to reductions in heart disease.

As a third limitation, we assumed the absence of dynastic effects, an assumption that is broken when parental genes associate with parental behaviours that directly cause a health outcome in the child.[48] For example, parents with a genetic predisposition towards higher education may choose to feed their children a better diet. However, parental educational attainment has been shown to be a poor predictor of conventional cardiovascular risk factors in children.[49] Fourthly, our observational and genetic data originate predominantly from samples of European origin in high income countries. We are thus unable to generalise these estimates to other populations, particularly to low income countries where cardiovascular diseases are less common. However, it may well be expected that socioeconomic factors mirror the pattern seen for other cardiovascular risk factors, whereby similar effects are typically seen across the world. For example, in the INTERHEART study, regional heterogeneity in the magnitude of associations was just as large for some conventional cardiovascular risk factors (eg, hypertension I^2=85%, obesity I^2=92%),[50] as it was for some psychosocial risk factors (eg, depression I^2=85%, general stress I^2=79%).[51] Fifthly, we do not know whether increasing education for the people with the least education will be as cardioprotective as increasing education for those with above average education. Nonetheless, a scenario of dose-response across the broad educational gradient is compatible with, firstly, the linear relation seen in the observational data. Secondly, it is also compatible with the concordance of findings from our study (which measures the average effect across the entire population) alongside the findings from studies of raising the school leaving age (which measure the effect among those with least education only).

Potential mechanisms

The mechanisms that might mediate the association between education and CHD remain relatively

unknown. Traditional observational associations have estimated that the association between education and CHD attenuates by around 30-45% after statistical adjustment for health behaviours and conventional cardiovascular risk factors (including smoking, blood pressure, and cholesterol); however, measurement error in such analyses can underestimate their mediating effect. This suggests that these factors could account for perhaps half of the association between education and CHD.[2 52] Our study found genetic predisposition towards longer education to associate with improved smoking, body mass index, and blood lipid profiles (with some borderline results for blood pressure and risk of diabetes). The degree of mediation should now be formally assessed with more extensive methods—for example, by applying two step mendelian randomisation.[53 54] If conventional risk factors do not completely account for the mechanism between education and CHD, then additional mechanistic hypotheses for investigations are needed. These could include education leading to improved use of healthcare services (from better health knowledge or fewer financial barriers to accessing care) or better job prospects, income, material conditions, social ranking and/or diet, all factors associated with education and CHD, many of which might be amenable to intervention.[4]

What our study adds

After exposure to a socioeconomic factor, there is often a long latency period before the occurrence of common diseases (in this example, around 50 years). Consequently, this line of research is not particularly amenable to randomised controlled trials, which would otherwise settle questions of causality. This does not mean that these associations are less worthy of investigation, particularly as large point estimates open up the possibility of potentially large public health gains. The solution is to triangulate evidence from multiple study designs, each with its own strengths and weaknesses. The limited studies to date have suggested that a causal effect between socioeconomic exposures and all cause mortality is more likely than not to exist. Our study adds to this evidence by using an entirely new technique, which also suggests that a causal effect is more likely than not to exist between education and CHD.

Implications for researchers

The main question for future research is "What mechanisms account for the strong association seen between genetic predisposition towards longer education and substantially lower risk of CHD?" Were it to be found that a health behaviour (such as diet) is an important mediator, then interventions on diet could become the cornerstone of policies designed to reduce health inequalities.

More molecular research is needed to delineate the mechanism, pleiotropic or not, through which these 162 education SNPs associate with cardiac outcomes. This could elucidate new causal mechanisms for CHD which, in turn, could lead to insights for potential drug discovery.

Implications for clinicians and policymakers

Although uncertainty remains around the precise function of each of the 162 SNPs, their degree of pleiotropy with cardiac traits, and the mechanisms by which these genetic variants exert their cardioprotective influence, conclusions can still be drawn from the current body of evidence. Firstly, policies that increase education probably lead to non-health benefits, such as increased economic productivity, higher voter turnout, better governance, and improved life satisfaction.[55 56] Secondly, very little evidence exists to suggest that increasing education might subsequently harm health or wellbeing. Thirdly, although rigorous scientific debate needs to continue on the health consequences of increasing education, the current balance of opinion seems to weigh towards the side on which increasing education will probably improve a range of health outcomes (either to a smaller or larger degree). Little discussion has taken place about how to increase education in a manner that is practical, acceptable, affordable, and sustainable. Although our data make no claims on this, we note that interventions should be accompanied by careful monitoring for unforeseen side effects, especially in those people who may not thrive when forced into extended educational settings, which may otherwise aggravate health inequalities. To briefly begin this discussion, one can imagine a range of policies by analogy to how clinicians, public health practitioners, and policymakers encourage patients to stop smoking: by raising awareness (for example, mass marketing campaigns, personalised letters, or individual counselling), convenience of access (for example, changing the geographical dispersion of educational establishments or opportunities for flexible education), and/or finance (for example, tuition fees, accommodation costs, or stipends). One can also consider complementing some of these population level policies with individual level interventions (for example, advising adolescents on whether to pursue higher education).

Conclusion

Our mendelian randomisation analyses found genetic support for the hypothesis that longer education plays a causal role in lowering the risk of coronary heart disease. Although completely ruling out possible pleiotropic effects is difficult, the sensitivity tests available to us gave little evidence that these could have driven our findings. In conjunction with the results from other study designs, increasing education is likely to lead to health benefits.

AUTHOR AFFILIATIONS

[1]Department of Epidemiology and Public Health, University College London, London, UK
[2]Department of Internal Medicine, Lausanne University Hospital, Lausanne, Switzerland
[3]Department of Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[4]Centre for Environmental Health Monitoring, National Institute of Public Health, Prague, Czech Republic

doi: 10.1136/bmj.j3542 | *BMJ* 2017;358:j3542 | the bmj

[5]Chair of Epidemiology and Population Studies, Institute of Public Health, Faculty of Health Sciences, Jagiellonian University Medical College, Krakow, Poland

[6]Institute of Cardiology, Lithuanian University of Health Sciences, Kaunas, Lithuania

[7]Research Institute of Internal and Preventive Medicine, Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

[8]Novosibirsk State Medical University, Novosibirsk, Russia

[9]Postgraduate Programme in Epidemiology, Federal University of Pelotas, Pelotas, Brazil

[10]Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

[11]Estonian Genome Center, University of Tartu, Tartu, Estonia

[12]Research Center in Epidemiology and Preventive Medicine, University of Insubria, Varese, Italy

[13]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

[14]School of Social and Community Medicine, University of Bristol, Bristol, UK

[15]Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, Big Data Institute, University of Oxford, Oxford, UK

[16]Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, UK

[17]National Institute for Health Research Oxford Biomedical Research Centre, Oxford University Hospital, Oxford, UK

We are grateful to the eight GWAS consortia (especially SSGAC and CARDIoGRAMplusC4D) for publicly sharing the genetic data we used in our causal analysis, to Reedik Mägi for assistance on LD score regression, and to Daniel J Benjamin for comments on an earlier draft of this article.

Contributors: TT and JV contributed equally to this paper and are joint first authors. TT had the idea for the study. AO obtained the genetic data. HP, AP, RK, AP, AT, SM, and KF obtained the observational data. MVH, GDS, JB, FPH, TP, and GV developed the study methods. JV, TT, FPH, and MVH did the analysis. All authors were involved in interpreting the data. TT, JV, and MVH wrote the first draft of the manuscript, and all authors critically revised it. TT, JV, and MVH are the guarantors.

Funding: TT is funded by a Wellcome Trust fellowship (106554/Z/14/Z); JV is supported by the Swiss National Science Foundation (P2LAP3_155086); the HAPIEE study is supported by the Wellcome Trust (064947/Z/01/Z, WT081081), the US National Institute on Aging (1R01 AG23522), the MacArthur Foundation (Health and Social Upheaval network), and the Russian Science Foundation (14-45-00030); the MORGAM Project was supported by the European Union's Seventh Framework Programme (HEALTH-F3-2010-242244, HEALTH-F2-2011-278913); GDS works in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC_UU_12013/1); JB is supported by a Medical Research Council (MRC) methodology research fellowship (grant MR/N501906/1); AO is supported by a European Research Council consolidator grant (647648 EdGe); MVH is supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre and works in the MRC Population Health Research Unit at the University of Oxford that receives funding from the UK MRC. The funders had no role in the study design, data collection, analysis, interpretation, or writing, nor in the decision to submit the article for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: support for the submitted work as detailed above; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Participants gave informed consent for data sharing, as described in each of the discovery genome-wide association studies. Additional ethical approval was not needed for this study.

Transparency: The lead authors affirm that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Data sharing: All of the summary level data used are available for instant download at the public repositories listed in table 1. The statistical code is available from the corresponding authors at t.tillmann@ucl.ac.uk and julien.vaucher@chuv.ch.

1 Hinkle LEJr, Whitney LH, Lehman EW, et al. Occupation, education, and coronary heart disease. Risk is influenced more by education and background than by occupational experiences, in the Bell System. *Science* 1968;161:238-46. doi:10.1126/science.161.3838.238
2 Manrique-Garcia E, Sidorchuk A, Hallqvist J, Moradi T. Socioeconomic position and incidence of acute myocardial infarction: a meta-analysis. *J Epidemiol Community Health* 2011;65:301-9. doi:10.1136/jech.2009.104075
3 Veronesi G, Ferrario MM, Kuulasmaa K, et al. Educational class inequalities in the incidence of coronary heart disease in Europe. *Heart* 2016;102:958-65. doi:10.1136/heartjnl-2015-308909
4 Davey Smith G, Hart C, Hole D, et al. Education and occupational social class: which is the more important indicator of mortality risk? *J Epidemiol Community Health* 1998;52:153-60. doi:10.1136/jech.52.3.153
5 Clark D, Royer H. The effect of education on adult mortality and health: evidence from Britain. *Am Econ Rev* 2013;103:2087-120. doi:10.1257/aer.103.6.2087.
6 Næss O, Hoff DA, Lawlor D, Mortensen LH. Education and adult cause-specific mortality--examining the impact of family factors shared by 871 367 Norwegian siblings. *Int J Epidemiol* 2012;41:1683-91, author reply 1691-3. doi:10.1093/ije/dys143
7 Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1-22. doi:10.1093/ije/dyg070
8 Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ* 2005;330:1076-9. doi:10.1136/bmj.330.7499.1076
9 Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016;40:304-14. doi:10.1002/gepi.21965
10 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512-25. doi:10.1093/ije/dyv080
11 Burgess S, Dudbridge F, Thompson SG. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". *Am J Epidemiol* 2015;181:290-1. doi:10.1093/aje/kwv017
12 Swerdlow DI, Holmes MV, Kuchenbaecker K, et al. INterleukin-6 Receptor Mendelian Randomisation Analysis (IL6R MR) Consortium. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* 2012;379:1214-24. doi:10.1016/S0140-6736(12)60110-X
13 Holmes MV, Dale CE, Zuccolo L, et al. InterAct Consortium. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ* 2014;349:g4164. doi:10.1136/bmj.g4164
14 Nguyen TT, Tchetgen Tchetgen EJ, Kawachi I, et al. Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk. *Ann Epidemiol* 2016;26:71-6.e1, 3. doi:10.1016/j.annepidem.2015.10.006
15 Cuellar-Partida G, Lu Y, Kho PF, et al. Assessing the Genetic Predisposition of Education on Myopia: A Mendelian Randomization Study. *Genet Epidemiol* 2016;40:66-72. doi:10.1002/gepi.21936
16 Okbay A, Beauchamp JP, Fontana MA, et al. LifeLines Cohort Study. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016;533:539-42. doi:10.1038/nature17671
17 Deloukas P, Kanoni S, Willenborg C, et al. CARDIoGRAMplusC4D Consortium. DIAGRAM Consortium. CARDIOGENICS Consortium. MuTHER Consortium. Wellcome Trust Case Control Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* 2013;45:25-33. doi:10.1038/ng.2480
18 Peasey A, Bobak M, Kubinova R, et al. Determinants of cardiovascular disease and other non-communicable diseases in Central and Eastern Europe: rationale and design of the HAPIEE study. *BMC Public Health* 2006;6:255. doi:10.1186/1471-2458-6-255

19 Evans A, Salomaa V, Kulathinal S, et al. MORGAM Project. MORGAM (an international pooling of cardiovascular cohorts). *Int J Epidemiol* 2005;34:21-7. doi:10.1093/ije/dyh327
20 Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010;42:441-7. doi:10.1038/ng.571
21 Wain LV, Verwoert GC, O'Reilly PF, et al. LifeLines Cohort Study, EchoGen consortium, AortaGen Consortium, CHARGE Consortium Heart Failure Working Group, KidneyGen consortium, CKDGen consortium, Cardiogenics consortium, CardioGram. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* 2011;43:1005-11. doi:10.1038/ng.922
22 Willer CJ, Schmidt EM, Sengupta S, et al. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;45:1274-83. doi:10.1038/ng.2797
23 Morris AP, Voight BF, Teslovich TM, et al. Wellcome Trust Case Control Consortium, Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Asian Genetic Epidemiology Network–Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012;44:981-90. doi:10.1038/ng.2383
24 Dupuis J, Langenberg C, Prokopenko I, et al. DIAGRAM Consortium, GIANT Consortium, Global BPgen Consortium, Anders Hamsten on behalf of Procardis Consortium, MAGIC investigators. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010;42:105-16. doi:10.1038/ng.520
25 Wood AR, Esko T, Yang J, et al. Electronic Medical Records and Genomics (eMEMERGEGE) Consortium, MIGen Consortium, PAGEGE Consortium, LifeLines Cohort Study. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014;46:1173-86. doi:10.1038/ng.3097
26 Kitagawa EM, Hauser PM. *Differential mortality in the United States: a study in socio-economic epidemiology*. 1st ed. Harvard University Press, 1973doi:10.4159/harvard.9780674188471.
27 Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomisation via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017; [forthcoming] doi:10.1093/ije/dyx102.
28 Bulik-Sullivan B, Finucane HK, Anttila V, et al. ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015;47:1236-41. doi:10.1038/ng.3406
29 Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol* 2016;45:1961-74.
30 Hingorani A, Humphries S. Nature's randomised trials. *Lancet* 2005;366:1906-8. doi:10.1016/S0140-6736(05)67767-7
31 van Kippersluis H, O'Donnell O, van Doorslaer E. Long run returns to education: does schooling lead to an extended old age? *J Hum Resour* 2009;4:1-33.
32 Davies NM, Dickson M, Davey Smith G, van den Berg G, Windmeijer F. The causal effects of education on health, mortality, cognition, well-being, and income in the UK Biobank. *bioRxiv* 2016. doi:https://doi.org/10.1101/074815.
33 Lleras-Muney A. The relationship between education and adult mortality in the United States. *Rev Econ Stud* 2005;72:189-221. doi:10.1111/0034-6527.00329.
34 Mazumder B. Does education improve health? A reexamination of the evidence from compulsory schooling laws. *Econ Perspect* 2008;33:2-16.
35 Lager AC, Torssander J. Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. *Proc Natl Acad Sci U S A* 2012;109:8461-6. doi:10.1073/pnas.1105839109
36 Madsen M, Andersen AM, Christensen K, Andersen PK, Osler M. Does educational status impact adult mortality in Denmark? A twin approach. *Am J Epidemiol* 2010;172:225-34. doi:10.1093/aje/kwq072
37 Madsen M, Andersen PK, Gerster M, Andersen A-MN, Christensen K, Osler M. Are the educational differences in incidence of cardiovascular disease explained by underlying familial factors? A twin study. *Soc Sci Med* 2014;118:182-90. doi:10.1016/j.socscimed.2014.04.016
38 Lundborg P, Lyttkens CH, Nystedt P. The Effect of Schooling on Mortality: New Evidence From 50,000 Swedish Twins. *Demography* 2016;53:1135-68. doi:10.1007/s13524-016-0489-3
39 Næss O, Hoff DA, Lawlor D, Mortensen LH. Education and adult cause-specific mortality--examining the impact of family factors shared by 871 367 Norwegian siblings. *Int J Epidemiol* 2012;41:1683-91, author reply 1691-3. doi:10.1093/ije/dys143
40 Søndergaard G, Mortensen LH, Nybo Andersen AM, et al. Does shared family background influence the impact of educational differences on early mortality? *Am J Epidemiol* 2012;176:675-83. doi:10.1093/aje/kws230
41 Marioni RE, Ritchie SJ, Joshi PK, et al. Social Science Genetic Association Consortium. Genetic variants linked to education predict longevity. *Proc Natl Acad Sci U S A* 2016;113:13366-71. doi:10.1073/pnas.1605334113
42 Nelson CP, Hamby SE, Saleheen DC, et al. ARDIoGRAM+C4D Consortium. Genetically determined height and coronary artery disease. *N Engl J Med* 2015;372:1608-18. doi:10.1056/NEJMoa1404881
43 Arden R, Luciano M, Deary IJ, et al. The association between intelligence and lifespan is mostly genetic. *Int J Epidemiol* 2016;45:178-85. doi:10.1093/ije/dyv112
44 Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;45:1866-86.
45 de Walque D. Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education. *J Health Econ* 2007;26:877-95. doi:10.1016/j.jhealeco.2006.12.005
46 Swerdlow DI, Preiss D, Kuchenbaecker KB, et al. DIAGRAM Consortium, MAGIC Consortium, InterAct Consortium. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. *Lancet* 2015;385:351-61. doi:10.1016/S0140-6736(14)61183-1
47 Ference BA, Yoo W, Alesh I, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. *J Am Coll Cardiol* 2012;60:2631-9. doi:10.1016/j.jacc.2012.09.017
48 Fletcher JM. The promise and pitfalls of combining genetic and economic research. *Health Econ* 2011;20:889-92. doi:10.1002/hec.1745
49 Kvaavik E, Glymour M, Klepp KI, Tell GS, Batty GD. Parental education as a predictor of offspring behavioural and physiological cardiovascular disease risk factors. *Eur J Public Health* 2012;22:544-50. doi:10.1093/eurpub/ckr106
50 Yusuf S, Hawken S, Ounpuu S, et al. INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364:937-52. doi:10.1016/S0140-6736(04)17018-9
51 Rosengren A, Hawken S, Ounpuu S, et al. INTERHEART investigators. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364:953-62. doi:10.1016/S0140-6736(04)17019-0
52 Stringhini S, Sabia S, Shipley M, et al. Association of socioeconomic position with health behaviors and mortality. *JAMA* 2010;303:1159-66. doi:10.1001/jama.2010.297
53 Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. *Hum Mol Genet* 2016;25(R2):R149-56. doi:10.1093/hmg/ddw197
54 Varbo A, Benn M, Smith GD, Timpson NJ, Tybjaerg-Hansen A, Nordestgaard BG. Remnant cholesterol, low-density lipoprotein cholesterol, and blood pressure as mediators from obesity to ischemic heart disease. *Circ Res* 2015;116:665-73. doi:10.1161/CIRCRESAHA.116.304846
55 Botero J, Ponce A, Shleifer A. Education, Complaints, and Accountability. *J Law Econ* 2013;56:959-96doi:10.1086/674133.
56 Oreopoulos P, Salvanes KG. Priceless: The Nonpecuniary Benefits of Schooling. *J Econ Perspect* 2011;25:159-84doi:10.1257/jep.25.1.159.

Supplementary materials: Analytical dataset

A3.2. Mediation and international differences

PLOS | MEDICINE

RESEARCH ARTICLE

Psychosocial and socioeconomic determinants of cardiovascular mortality in Eastern Europe: A multicentre prospective cohort study

Taavi Tillmann[1]*, Hynek Pikhart[1], Anne Peasey[1], Ruzena Kubinova[2], Andrzej Pajak[3], Abdonas Tamosiunas[4], Sofia Malyutina[5,6], Andrew Steptoe[1], Mika Kivimäki[1], Michael Marmot[1], Martin Bobak[1]

1 Department of Epidemiology & Public Health, University College London, London, United Kingdom, **2** Centre for Environmental Health Monitoring, National Institute of Public Health, Prague, Czech Republic, **3** Chair of Epidemiology and Population Studies, Institute of Public Health, Faculty of Health Sciences, Jagiellonian University, Krakow, Poland, **4** Institute of Cardiology, Lithuanian University of Health Sciences, Kaunas, Lithuania, **5** Research Institute of Internal and Preventive Medicine, Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, **6** Novosibirsk State Medical University, Novosibirsk, Russia

* t.tillmann@ucl.ac.uk

OPEN ACCESS

Citation: Tillmann T, Pikhart H, Peasey A, Kubinova R, Pajak A, Tamosiunas A, et al. (2017) Psychosocial and socioeconomic determinants of cardiovascular mortality in Eastern Europe: A multicentre prospective cohort study. PLoS Med 14(12): e1002459. https://doi.org/10.1371/journal.pmed.1002459

Academic Editor: Kazem Rahimi, University of Oxford, UNITED KINGDOM

Received: April 14, 2017

Accepted: October 23, 2017

Published: December 6, 2017

Data Availability Statement: Data are from the HAPIEE study, which are available on request from m.bobak@ucl.ac.uk and ethics@ucl.ac.uk, who will jointly seek approval by the HAPIEE Study Steering Committee and the Research Ethics Committee at UCL and participating centres. Public data deposition was not possible under Russian law.

Funding: This work was supported by The Wellcome Trust [106554/Z/14/Z to TT; 064947/Z/01/Z to MB, HP, APe, RK, APa, AT, SM, and MM;

Abstract

Background

Eastern European countries have some of the highest rates of cardiovascular disease (CVD) mortality, much of which cannot be adequately accounted for by conventional CVD risk factors. Psychosocial and socioeconomic factors may affect risk of CVD, but relatively few studies on this issue have been undertaken in Eastern Europe. We investigated whether various psychosocial factors are associated with CVD mortality independently from each other and whether they can help explain differences in CVD mortality between Eastern European populations.

Methods

Participants were from the Health, Alcohol and Psychological factors in Eastern Europe (HAPIEE) cohort study in Russia, Poland and the Czech Republic, including a total of 20,867 men and women aged 43–74 years and free of CVD at baseline examination during 2002–2005. Participants were followed-up for CVD mortality after linkage to national mortality registries for a median of 7.2 years.

Results

During the follow-up, 556 participants died from CVD. After mutual adjustment, six psychosocial and socioeconomic factors were associated with increased risk of CVD death: unemployment, low material amenities, depression, being single, infrequent contacts with friends or relatives. The hazard ratios [HRs] for these six factors ranged between 1.26 [95%

WT081081 to MB, HP, APe, RK, APa, AT, SM, and MM]; the US National Institute on Aging [1R01 AG23522 to MB, HP, RK, APe, APa, AT, SM, and MM]; the MacArthur Foundation [Health and Social Upheaval network to MB, RK, APa, AT, SM, and MM]; the Seventh Framework Programme of the European Commission [HEALTH-F3-2010-242244 to APe]; the Medical Research Council [K013351 to MK]; and the Russian Scientific Foundation [14-45-00030 to SM, MB, HP, APe, and TT]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: APa declares previous honoraria from Amgen and Sanofi, which are not related to this article. The other authors have declared that no competing interests exist.

Abbreviations: CI, confidence interval; CVD, cardiovascular disease; GFQ, graduated frequency questionnaire; HAPIEE, Health, Alcohol and Psychosocial factors In Eastern Europe; HR, hazard ratio; MONICA, Multinational MONItoring of trends and determinants in CArdiovascular disease; PAF, population attributable fraction.

confidence interval 1.14–1.40] and 1.81 [95% confidence interval 1.24–2.64], fully adjusted for each other, and conventional cardiovascular risk factors. Population-attributable fractions ranged from 8% [4%–13%] to 22% [11%–31%] for each factor, when measured on average across the three cohorts. However, the prevalence of psychosocial and socioeconomic risk factors and their HRs were similar between the three countries. Altogether, these factors could not explain why participants from Russia had higher CVD mortality when compared to participants from Poland/Czech Republic. Limitations of this study include measurement error that could lead to residual confounding; and the possibilities for reverse causation and/or unmeasured confounding from observational studies to lead to associations that are not causal in nature.

Conclusions

Six psychosocial and socioeconomic factors were associated with cardiovascular mortality, independent of each other. Differences in mortality between cohorts from Russia versus Poland or Check Republic remained unexplained.

Author summary

Why was this study done?

- Countries in Eastern Europe have unusually high rates of cardiovascular disease, but the causes of this remain unknown.
- Second, studies from around the world have found that people with less education, income, and/or wealth develop more heart disease, but it is unclear why this happens.

What did the researchers do and find?

- We examined data from 20,867 healthy middle-aged adults from Russia, Poland, and the Czech Republic, and asked them a range of questions about stress-related factors. After a median of 7.2 years, 556 of them had died of heart disease.
- We confirmed that people with conventional risk factors (e.g., smoking, high blood pressure, diabetes, physical inactivity, and obesity) were more likely to die from heart disease.
- We also saw that people who were unemployed, single, with less wealth, who rarely saw their friends and relatives, and who are more depressed were more likely to die from heart disease. These associations with heart disease were not explained by the conventional risk factors, and they seemed to be independent of one another.
- After considering country-specific differences in conventional and stress-related factors, we were not able to explain why the risk of cardiovascular death is twice as high in Russia when compared to Poland or the Czech Republic.

What do these findings mean?

- We still do not understand why life expectancy remains so low in many areas of Eastern Europe, such as Russia. The HAPIEE study suggests that alcohol, diet, and stress are unlikely to cause these international differences. New hypotheses are now required to investigate this further.
- There may be at least six stress-related risk factors for heart disease. Each of these could be further investigated to see if they improve clinical risk prediction models, if they are causal, and how they work.

Introduction

Psychosocial and socioeconomic risk factors, such as unemployment, low social support, and depression, are associated with increased risk of cardiovascular disease (CVD) [1, 2]. Meta-analyses suggest that the excess risk associated with some of these factors can be as high as that for hypertension or raised cholesterol [3–5]. According to the 2016 European guidelines for CVD prevention, these factors are potentially a useful target for intervention [6]. However, the evidence to support this suggestion remains limited.

One weakness in the current evidence base is that most data comes primarily from Western Europe and Northern America, making it unclear to what extent these findings are generalisable elsewhere. For example, Eastern European countries have some of the highest rates of CVD mortality in the world, most of which cannot be adequately accounted for by conventional CVD risk factors [7]. The high rates of CVD mortality in Eastern European countries has a large impact on life expectancy in these countries, which in 2015 was only 65 for men in Russia, compared to 67 for men in India or Cambodia [8]. The 2015 Global Burden of Disease study highlighted how the Eastern European region deviates strikingly from the otherwise tight correlation between life expectancy and socioeconomic development (see figure 10 from reference[9]). This deviation remains an important global health question that could inform development goals elsewhere.

It has been hypothesised that the transition from communism to capitalism (which began in 1989) in Eastern European countries could have exacerbated the influence of psychosocial hazards on CVD [10]. During the early 1990s, countries like Russia experienced a recession many times larger than the recent 2008 global recession. Together with vanishing social welfare, rising crime, political uncertainty, and changing cultural expectations, this period of massive social change coincided, in some countries, with large increases in suicide and CVD mortality, especially among single middle-aged men of low education [11, 12]. However, there have been no prospective studies investigating the role of other psychosocial and socioeconomic risk factors in predicting CVD in this region, such as unemployment, depression, social support, perceived control, material deprivation, material amenities, and loss of social status. Furthermore, literature from other countries is relatively sparse about the extent to which depression and social support might explain socioeconomic differences in CVD mortality.

The aim of this multicentre cohort study was to investigate the extent to which psychosocial and socioeconomic factors are associated with CVD mortality in three Eastern European populations. We examined whether these risk factors are independently associated with CVD even after adjusting for each others' effects and whether psychosocial and socioeconomic factors

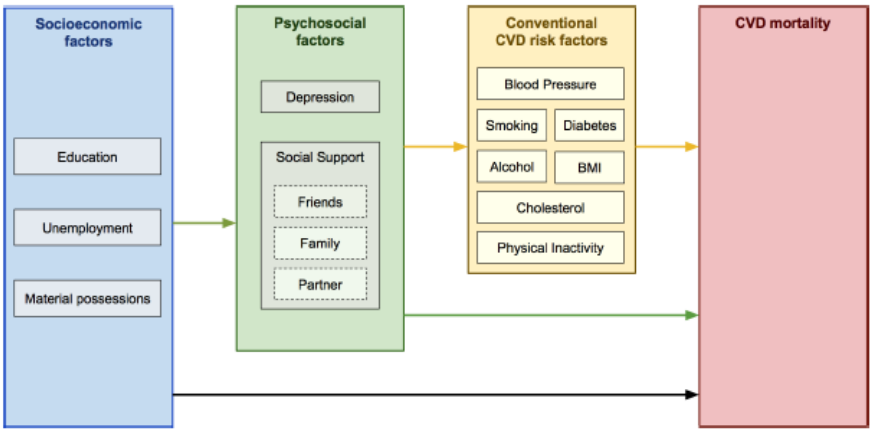


Fig 1. Theoretical direction of causal effects, from three sets of risk factors, to CVD mortality. BMI, body mass index; CVD, cardiovascular disease.

https://doi.org/10.1371/journal.pmed.1002459.g001

help to explain the high rates of CVD mortality in Russia when compared to similar populations with lower rates (Poland and the Czech Republic). In addition, we also explored whether depression and lack of social support might explain socioeconomic differences in CVD mortality (as measured by educational attainment, unemployment, or material possessions). These concepts are illustrated in our causal diagram (Fig 1).

Methods

Ethics statement

The study was approved by the University College London/University College London Hospital ethics committee and by the local ethics committee in each participating centre. All participants gave written informed consent.

Rationale

The Health, Alcohol and Psychosocial factors In Eastern Europe (HAPIEE) study is a multicentre prospective cohort study of urban populations living in Russia, Poland, and the Czech Republic. The rationale for study was described in a cohort description published in 2006 [13]. Briefly, the study was designed to investigate the role of nutritional-, psychosocial-, and alcohol-related factors in the incidence of common disease (both at the within-country level, as well as to account for international differences in rates of disease). Previous publications from this study have reported associations between mortality and nutrition [14], alcohol consumption [15], and socioeconomic status [16]. Nutrition and alcohol did not explain why participants from Russia had twice the mortality rate when compared with participants from Poland or the Czech Republic. The current publication therefore assesses, in a systematic fashion, several original objectives of the study. We do not use data from the Lithuanian arm of the HAPIEE study as its baseline questionnaire did not include all the psychosocial variables of interest.

Participants

Random population samples of 28,945 men and women aged 43–72 years at baseline during 2002–2005 were selected from population registers and electoral lists. The overall response rate was 59% (61% in Russia and Poland and 55% in the Czech Republic). We excluded 7,173 participants (25%) with a history of acute myocardial infarction, stroke, chronic heart disease, or angina, as well as those who scored positively on the Rose Angina Questionnaire.

Participants were linked to the national mortality registries using a national personal ID number in the Czech Republic and Poland. In Russia, linkage was done to the Novosibirsk Oblast regional mortality registry (which covers a land area larger than the Czech Republic) using surname, initials, and date of birth. Any potential inconsistencies were corrected manually, first by extracting the participant's address and second (in case of inconclusive linkage) by telephoning the participant's next of kin for verification. Nine hundred and five participants (4.2%) were lost to follow-up, primarily since they did not give consent to link their records with the national mortality registries, or due to migration (losses of 4%, 7%, and 1% in Czech, Polish, and Russian samples, respectively). This left an analytical sample of 20,867 participants. Follow-up was available to 31st December 2010 in Russia and Poland and 30th June 2013 in the Czech Republic (maximum follow up of 8.0, 8.9, and 11.3 years, respectively). The primary endpoint was CVD mortality (ICD-10 codes I00-99).

Of the participants who participated in wave 1, 66% also participants also participated in wave 2 of the HAPIEE study. This was collected in 2006–2008, i.e., after a mean interval of 3.6 (range 1 to 6) years. In this paper, data from wave 2 were only used to inform the predictor matrix when imputing missing data, and were not themselves used as exposures or outcomes.

Socioeconomic factors

At baseline, participants completed an extensive structured questionnaire, and underwent a standardised nurse examination in a clinic. Highest educational qualification was grouped into three categories: primary or less, secondary, and tertiary. Assuming that the average participants in these groups were separated by around three and six years of additional schooling, respectively, we modelled a linear relationship between these three categories. For self-reported economic activity, participants were classified into three groups: economically active; retired and no longer working; and currently unemployed. Participants were also asked about long-term unemployment, with four options (never, up to 3 months total, up to 3–12 months total, more than 12 months total); responses were dichotomised, comparing those unemployed for more than 12 months against the rest.

Possession of 10 material amenities (microwave, video recorder, colour television, washing machine, dishwasher, freezer, camcorder, satellite TV, telephone, and mobile phone; each coded 0/1) was used to derive a standardised continuous scale. For current material deprivation, three questions asked about how often the subjects had difficulties with paying for food, clothes, or bills. Each item was scored 0–3 to yield a total score of 0–12. This was modelled per one standard deviation greater material deprivation. All standardisation procedures used the mean and SD obtained from all three cohorts in combination (as opposed to standardising, so that the standardised mean in each country is zero). Similar to current material deprivation, another scale of early-life material deprivation asked about the same items, this time with participants retrospectively recalling their childhood.

In addition, participants were asked whether the *'Changes since 1989 have been good or bad for your general social position?'* Original responses in five categories (very good, good, no change, bad, very bad) were regrouped into three categories.

Psychosocial factors

Social support is typically measured with a multifaceted instrument, covering aspects of friends, family, social clubs, and marital status. Given the larger power of our study, we did not combine these separate domains. For marital status, we compared the 'married/cohabiting' reference group against 'divorced/widowed', or 'single'. *'Are you a member of a club/organization'* was kept binary. '*How often are you in contact with relatives outside of your household?*' originally had six options: '*several times a week*'; '*about once a week*'; '*several times a month*'; '*about once a month*'; '*less than once a month*'; '*I don't have any relatives*'. We dichotomised this at the *'at least monthly'* threshold. A separate question, '*How often do you visit friends outside of your household?*', was handled similarly.

Depressive symptoms were assessed with the CESD-20 questionnaire (range 0–60).[17] A binary trait for depressive symptoms was defined as CESD-20 score $\geq$ 16. Perceived control was assessed with a scale developed by the MacArthur programme on midlife development [18]. The subscale 'control over life' (range 0–40) was used as a standardised continuous variable.

Conventional cardiovascular risk factors and other covariates

Covariates measured at baseline included country, age, gender, and 11 conventional CVD risk factors: smoking status (five categories of never smoker, ex-smoker; 1–10, 11–20, and 21+ cigarettes a day), diabetes; and physical activity (dichotomised at 2.5 hours a week) were determined by self-report. Clinical examination [13] determined systolic blood pressure (modelled linearly from 115 mmHg onwards [19]); as well as BMI and total and HDL serum cholesterol (whereby all three were modelled with linear and square terms after centring at 23 kg/m^2, 6 mmol/L, and 1.5 mmol/L, respectively). Alcohol intake in the last 12 months was self-reported using the graduated frequency questionnaire (GFQ) [20]. The United Kingdom Chief Medical Officer's alcohol guideline advises men and women to consume less than 14 units of alcohol (equivalent to 112 grams of ethanol) per week. Participants were categorised as either non-drinkers, drinking within these guidelines, drinking up to twice the guideline limit, or drinking more than twice the guideline limit. For completeness, we included three additional alcohol-related covariates. Frequency of alcohol consumption was dichotomised at the once a week threshold. A pattern of binge drinking was defined if men/women reported consuming more than 100 g/60 g of ethanol in one episode at least monthly. The CAGE questionnaire was used to evaluate symptoms of problems with alcohol and was dichotomised at 2 or above.

Statistical analyses

Missing data. Between 0%–12% of the data was missing for each variable (Tables 1 and S1). This was imputed from 10 multiple imputation models that included vital status, follow-up time, and all covariates. The predictor matrix additionally included key variables that were associated with values for missing variables (where these were known), as well as their absence (where these were unknown), taken from either the baseline survey or a follow-up survey done three years later on the same participants. These predictors were further self-reported details about the time that participants spent on sports, games, and hiking; the tendency to rely only on self (as opposed to others); overcommitment at work; amount of trust in the local area; perceptions that money influences health; perceptions that others treat the participant unfairly; unemployment of others in the participant's household; hypercholesterolaemia; antihypertensive drug and statin usage; and subsequent depression in wave 2.

Main analysis. Cox regression was used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for associations between psychosocial factors and mortality endpoints,

Table 1. Baseline characteristics of the analytical sample.

	Total sample		CVD mortality		% missing and imputed
	n / mean	% / SD	n / mean	% / SD	
Participants	20,867	100%	556	2.7%	
Follow-up years (median, max)	7.2	11.3	4.4	10.0	
Age	57.2	7.03	62.0	6.20	0%
Male	9,700	46%	380	68%	0%
Conventional CVD risk factors					
Diabetes	1,554	7.4%	93	17%	0.1%
Smoking status:					0.6%
Never-smoker	9,941	48%	172	31%	
Occasional/past smoker	5,065	24%	133	24%	
Daily smoker, 1–10 cigarettes/day	2,074	9.9%	71	13%	
Daily smoker, 11–20 cigarettes/day	3,039	15%	142	25%	
Daily smoker, >20 cigarettes/day	747	3.6%	38	6.9%	
Blood pressure, systolic (mmHg)	138.9	22.5	153.2	25.8	9.2%
Total cholesterol (mmol/L)	5.95	1.25	6.06	1.44	10.5%
HDL cholesterol (mmol/L)	1.48	0.46	1.48	0.50	10.6%
Body mass index (kg/m^2)	28.0	4.85	28.1	5.80	9.2%
Physically inactive	1,666	8.0%	100	18%	1.1%
Alcohol intake:					1.2%
Nil	4,033	19%	144	26%	
Up to UK guidelines	12,389	59%	274	49%	
Exceeding UK guidelines (1–2x over)	2,314	11%	65	12%	
Exceeding UK guidelines (>2x over)	2,131	10%	74	13%	
Alcohol drinking frequency:					1.4%
Nondrinker	3,942	19%	142	26%	
< once/week	10,534	50%	218	39%	
≥ once/week	6,391	31%	196	35%	
Binge drinking (≥ once/month)	2,705	13%	103	19%	1.4%
Possible problem drinking (CAGE ≥ 2)	1,408	6.7%	79	13%	8.1%
Psychosocial factors					
Marital status:					0.2%
Married/cohabiting	15,713	75%	380	68%	
Divorced/widowed	4,257	20%	143	26%	
Single	897	4.3%	33	5.9%	
Social support					
Contacts relatives < once/month	4,966	24%	181	33%	0.4%
Contacts friends < once/month	7,554	36%	224	40%	0.6%
Not a member of a club	17,487	84%	491	88%	0.6%
Depression symptoms (possible case)	4,612	22%	161	29%	11.9%
Low perceived control (SD scale)	0.00	1.01	0.29	1.08	1.5%
Socioeconomic factors					
Education					0.2%
Tertiary	5,263	25%	90	16%	
Secondary	13,459	65%	354	64%	
Primary	2,143	10%	112	20%	
Material possessions					
Low amenities, current (SD scale)	0.00	1.00	0.53	1.00	2.6%

(*Continued*)

Table 1. (*Continued*)

	Total sample		CVD mortality		% missing and imputed
	n / mean	% / SD	n / mean	% / SD	
Low amenities, early life (SD scale)	0.00	1.01	0.38	0.96	2.9%
Deprivation, current (SD scale)	0.00	1.02	0.22	1.16	0.9%
Deprivation, early life (SD scale)	0.00	0.99	0.05	1.11	1.0%
Unemployment, current	897	4.3%	28	5.0%	0.4%
Unemployment, long term	1,690	8.1%	44	7.8%	1.8%
Change in status since 1989:					0.9%
Improved	5,071	24%	103	19%	
Stayed the same	10,079	48%	261	47%	
Declined	5,718	27%	192	33%	

The sample shown is multiply imputed, combined across three countries and two genders. The final column shows the amount of missing/imputed data among the total sample (20,867 participants).

Abbreviations: HDL, high density lipoprotein; SD, standard deviation.

https://doi.org/10.1371/journal.pmed.1002459.t001

using follow-up time as the time scale. Data were pooled across three cohorts and both genders. Three models were created with increasing levels of adjustment. Model 1 was adjusted for age, sex, and country. International differences in cardiovascular mortality in this region are known to be much larger in men than women. To examine potential causes of this, we looked at whether gender and/or being placed in Russia modifies the association between psychosocial/socioeconomic risk factors and CVD mortality. We included interactions that met a Bonferroni-adjusted p-value threshold of $0.05/14 = 0.0036$. In model 2, we additionally adjusted for 11 conventional CVD risk factors for two reasons. First, as an indication of how much of the hazard from psychosocial/socioeconomic exposures might plausibly be mediated indirectly via conventional risk factors, and, second, as an indication of their direct effect size, through pathways not measured in this study. Model 3 was additionally adjusted for the six psychosocial and socioeconomic factors that are associated with CVD mortality (at $p<0.05$), following backwards-stepwise elimination from a larger model that began with all 14 candidate psychosocial factors. This approach was applied for three reasons: first, to help address statistical collinearity among the 14 exposures of interest; second, to identify a shortlist of top risk factors that might aetiologically lie on mutually exclusive pathways; and third, to identify a shortlist of top predictors that might be prioritised in future clinical care applications, such as novel risk prediction models. The proportional hazards assumption was tested using Schoenfeld residuals and checked graphically with log-log plots, with no evidence of its violation.

In all three models, we assumed no differences in the size of the risk factor–mortality associations between participants located in Poland, the Czech Republic, and Russia. There was insufficient power to explore heterogeneity between Poland and the Czech Republic, but we examined effect modification between participants from Russian versus Czech/Polish cohorts. Additional sensitivity analyses looked for effect modification by gender, repeating the main analysis after excluding imputed data, excluding participants with less than two years of follow-up (to reduce reverse causation bias), and using a similar 8-year follow-up for all three countries. In addition, we examined associations with all-cause mortality as the outcome instead of cardiovascular mortality.

Mediation analysis. Simple and complex models were compared with each other to evaluate the degree of attenuation and thereby infer approximate degree of mediation, using the

formula:

$$\text{Attenuation, or amount mediated} = [\log(\text{Fully adjusted HR}) - \log(\text{Crude HR})] \,/\, [\log(\text{Crude HR})].$$

For example, if the association between education and mortality is HR = 1.45 in model 1, and this attenuates to HR = 1.20 in model 2, then one can infer that the conventional cardiovascular risk factors (which are additionally included in model 2 but not in model 1) might account and potentially mediate around half of the pathway from education to mortality. Notably, this is a relatively crude method that is prone to differential measurement error as well as model mis-specification. Accordingly, we used this method only to provide very approximate and qualitative inferences about whether putative mediators are likely to play a small or large role, respectively.

Population Attributable Fraction. The Population Attributable Fraction denotes the proportion of mortality that could be prevented if the entire population were not exposed to a given risk factor (and assuming a causal relationship between exposure and outcome) [21]. In our study, it was calculated by fitting model 2 to the first set of imputed data and then using the *punafcc* package in STATA 14. As this package is unable to handle continuous risk factors, the continuous variable *material amenities* was dichotomised into two halves using the median as a cut off. For education, we calculated attribution if everybody without tertiary education would attain tertiary education. BMI was dichotomised as obese (BMI > 30 kg/m^2) or not. This study is reported as per STROBE guidelines (S1 Checklist).

Analysis formulation. While the overall analyses we report correspond with the prespecified aims of the study, we did not have a detailed analysis plan prespecified. Strategic decisions about which analytic approach to use were mostly made before the analysis by applying methods that we have previously applied in other publications. For example, our mediation analysis has previously been applied to other publications from the HAPIEE and Whitehall II cohorts. [14, 22, 23] Our evaluation of population-attributable risk reflects the method we have previously applied for the IPD-Work consortium.[24] The a priori variables we included in models 1 and 2 reflect the approach we have taken with the MORGAM consortium [2]. In contrast to our previous publications, the current analysis investigates multiple collinear exposures. Our decision to use a stepwise approach to inform variable selection in model 3 was inspired by the use of this technique in the derivation of clinical risk prediction models [25]. Alcohol consumption was not initially included in the analysis (since our previous publication found that alcohol did not explain differences in rates of disease between cohorts), but alcohol-related covariates were later included in the analysis, at the request of peer reviewers.

Results

Baseline characteristics of participants in the analytical sample are shown in Table 1. There were 556 deaths from CVD (249 out of 6,923 participants in Russia, 134 out of 7,039 in Poland, and 173 out of 6,905 in the Czech Republic). Participants who subsequently died of CVD had higher levels of most risk factors when compared to those who did not die of CVD. There was an interaction between sex and country: men in Russia had higher HRs than expected based on the sum of the 'male' and the 'Russian' indicator variables alone (HR for the interaction term = 1.77 [1.23–2.55], $p = 0.002$). This term was kept in all subsequent models.

Models 1 and 2

As expected, conventional CVD risk factors were associated with CVD mortality (Fig 2, S4 Table). Associations between 14 psychosocial and socioeconomic exposures and CVD mortality are shown in S5 Table. In model 1, 13 out of 14 psychosocial factors tested were

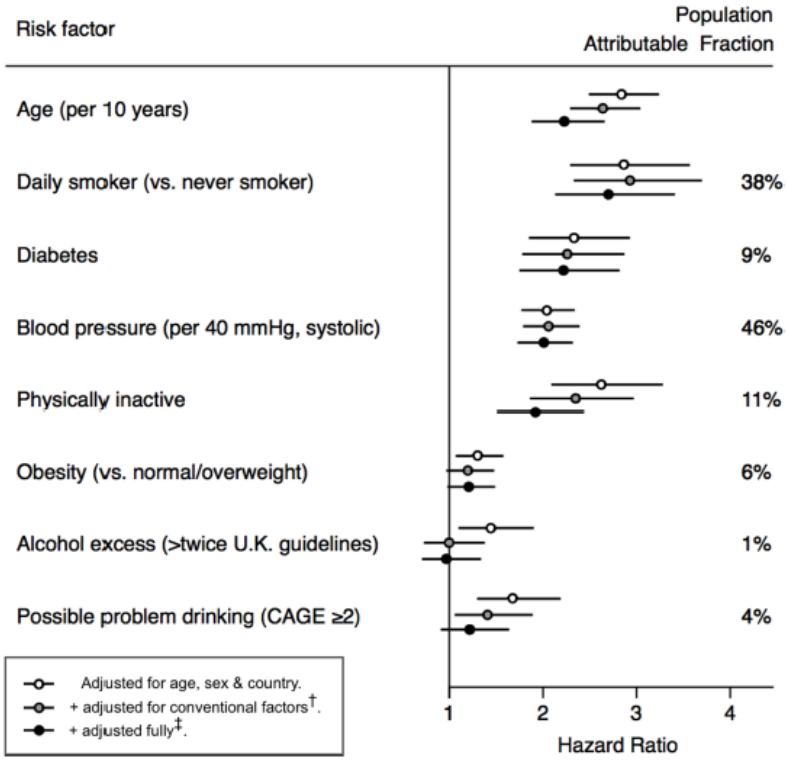


Fig 2. Associations of conventional cardiovascular risk factors with cardiovascular mortality.
†Conventional factors = diabetes, smoking, blood pressure, cholesterol, HDL, BMI, physical activity, alcohol.
‡Full adjustment = conventional factors + material possessions, depression, contacting relatives, contacting friends, friends*gender interaction, marital status, unemployment.

https://doi.org/10.1371/journal.pmed.1002459.g002

associated with CVD mortality, with HRs ranging from 2.96 (1.97–4.46, $p < 0.0001$) for current unemployment to HR = 1.14 (1.05–1.23, $p = 0.012$) per one standard deviation increase in early life deprivation. Twelve associations remained significant after adjustment for eleven conventional CVD risk factors in model 2—this attenuated the remaining HRs by around a quarter.

Infrequent contact with friends was associated with outcomes more strongly in women than in men (Model 1 interaction with sex HR = 1.83 [1.26–2.66], $p = 0.002$, thus satisfying Bonferroni criteria). Model 3 therefore included both the conventional binary variable '*low friends*' (which was handled similarly to the other five psychosocial factors) as well as the interaction term '*gender* low friends*' (interactions were not used for the remaining five psychosocial and socioeconomic factors). We did not see any evidence of effect modification by country (S7 Table).

Model 3

Following full adjustment for other psychosocial and socioeconomic factors in model 3, six factors remained associated with the outcome ($0.00001 < p < 0.007$): *depression*, *low material amenities*, *current unemployment*, *infrequent contact with relatives*, *infrequent contact with friends (for female participants only)*, and *single marital status* (Fig 3). Following this full adjustment, the HR for education was largely attenuated in comparison with age-sex adjusted models. Consequently, we did not consider education to be one of the core socioeconomic variables in subsequent multivariate analyses. The PAFs in models adjusted for conventional risk factors ranged from around 8% [4%–13%] for infrequent contact with relatives to 22% [11%–31%] for low material amenities. Test of effect modification showed similar results in analyses stratified by gender or cohort (S6 and S7 Tables). Sensitivity analyses gave similar results when limiting follow-up time to eight years in all three countries, excluding those participants with less than two years of follow-up, excluding imputed data, when using all-cause mortality as the outcome (S8–S11 Tables), or when increasing the number of psychosocial/socioeconomic covariates (in model 3) from 6 to all 14 factors (model 4 from S5 Table). In the full model for all-cause mortality, evidence of independent associations additionally emerged for *no club membership* and *education* (S11 Table).

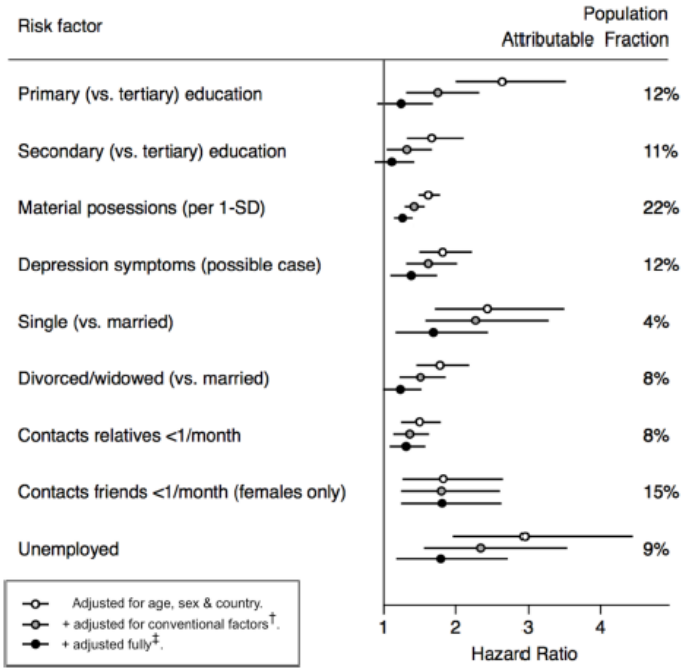


Fig 3. Associations of psychosocial and socioeconomic factors with cardiovascular mortality. SD, Standard Deviation

†Conventional factors = diabetes, smoking, blood pressure, cholesterol, HDL, BMI, physical activity, alcohol.

‡Full adjustment = conventional factors + material possessions, depression, contacting relatives, contacting friends, friends*gender interaction, marital status, unemployment.

https://doi.org/10.1371/journal.pmed.1002459.g003

Mediation analysis

We examined to what degree the HRs associated with the psychosocial and socioeconomic factors attenuated following various adjustments (Fig 4 and S12 Table). Adjustment for 11 conventional risk factors attenuated these by about one quarter. Additional adjustment for other psychosocial and socioeconomic factors attenuated these by an additional quarter. As two exceptions, the HRs for marital status and limited contact with friends and relatives did not attenuate to this level. This suggests very little overlap between the potential effects of these facets of social support, and that conventional risk factors may play a negligible role in mediating any of these effects. Overall, our data suggests that depression and social support constructs, if causal, may operate independently to socioeconomic constructs, potentially along separate mechanistic pathways.

As few studies have reported to what degree the associations between alcohol and mortality attenuate after adjustment for psychosocial factors, we additionally report these here. There was little evidence that total alcohol consumption or binge drinking was associated with CVD mortality in models adjusted for conventional risk factors. However, people scoring positive on the CAGE screening questionnaire for possible problems with alcohol had 41% greater risk

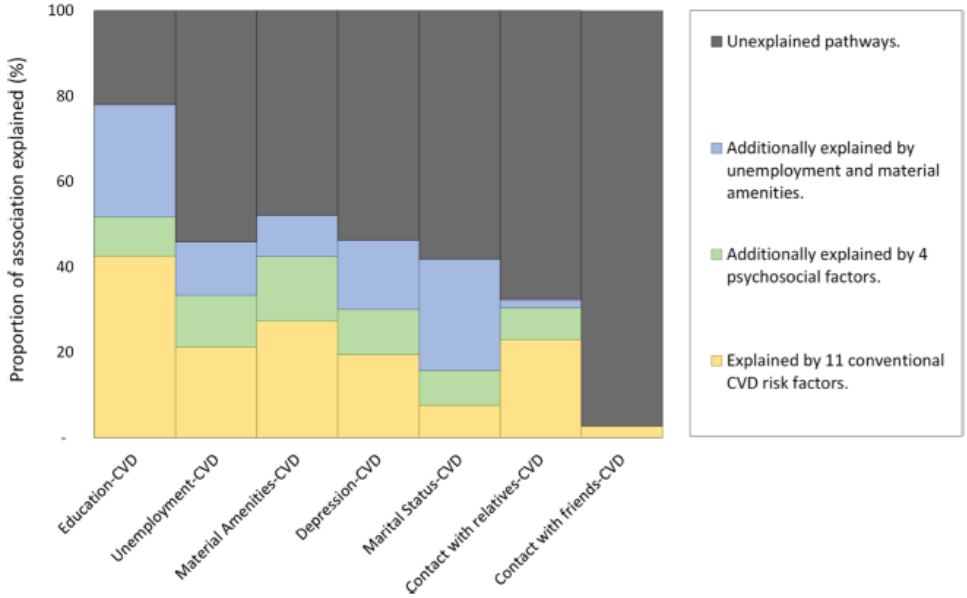


Fig 4. Attenuation among three socioeconomic (left side) and four psychosocial (right side) predictors of cardiovascular mortality. Created from four sequentially-adjusted models. The total height of each predictor on the y-axis (i.e., 100%) is equivalent to its association with cardiovascular mortality in model 1 (adjusted only for age, sex, and country). The yellow area represents the subsequent attenuation in association, following additional adjustment for 11 conventional CVD risk factors in model 2 (broadly corresponding to the yellow arrows in Fig 1). The green area represents the subsequent attenuation, following additional adjustment for four psychosocial variables (i.e., green arrow in Fig 1; coefficients shown in S12 Table). The blue area represents subsequent attenuation, following additional adjustment for two socioeconomic variables in model 3. The black area accounts for the unexplained part still present in model 3 (i.e., direct or nonmediated effects; black arrow in Fig 1). CVD, cardiovascular disease.

https://doi.org/10.1371/journal.pmed.1002459.g004

of CVD in models adjusted for age, sex, and country. This attenuated by around one half, following adjustment for six psychosocial and socioeconomic factors.

International differences

CVD mortality risk was substantially higher in the Russian cohort. In men, the age-adjusted HR for being in Russia versus Central Europe was 2.86 [2.31–3.54]; this excess risk was not reduced following adjustment for conventional CVD risk factors (HR = 2.78 [2.15–3.59]) or following additional adjustment for psychosocial and socioeconomic factors (HR = 2.77 [2.13–3.61]) (Fig 5). In women, the HR of being in Russia versus Central Europe was 1.59 [1.15–2.19]; this difference was exacerbated following adjustment for conventional factors (HR = 2.15 [1.45–3.18]) but returned to a level comparable to crude models following additional adjustment for psychosocial and socioeconomic factors (HR = 1.64 [1.09–2.46]) (S13 Table).

Discussion

This large prospective cohort study including adults from Russia, Poland, and the Czech Republic found independent associations between six psychosocial factors and subsequent cardiovascular mortality following full mutual adjustment. Depression and social support factors

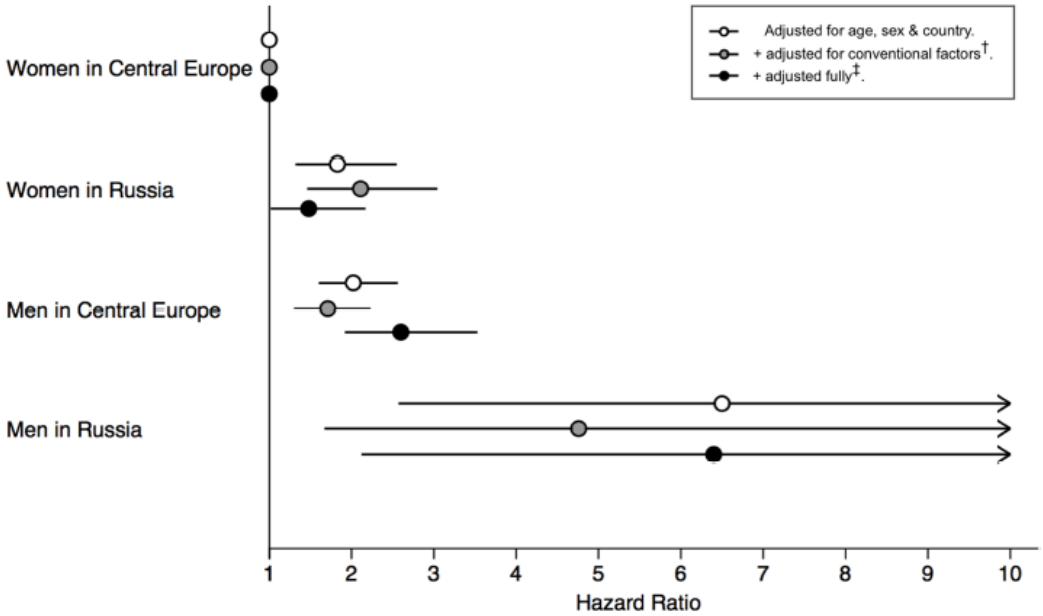


Fig 5. Increased risk of cardiovascular mortality from being male and/or being in Russia (versus being a female in Central Europe). Central Europe = Poland or Czech Republic. There were 109, 67, 198, and 175 events, respectively in the four groups shown (from top to bottom).
†Conventional factors = diabetes, smoking, blood pressure, cholesterol, HDL, BMI, physical activity, alcohol.
‡Full adjustment = conventional factors + material possessions, depression, contacting relatives, contacting friends, friends*gender interaction, marital status, unemployment.

https://doi.org/10.1371/journal.pmed.1002459.g005

did not substantially attenuate socioeconomic gradients in CVD mortality, suggesting that these mechanisms are more likely to be complementary, as opposed to being mediatory. Contrary to our expectations, the twofold higher risk of CVD mortality seen in participants from Russia (when compared to participants from Poland or the Czech Republic) was not reduced following adjustment for conventional, psychosocial, and socioeconomic factors.

Comparison with research in Western Europe and Northern America

Most of the associations we report are broadly consistent with prior studies, primarily from more affluent countries where all-cause mortality is often reported more commonly than our outcome of cardiovascular mortality.[2–5] However, we found current unemployment to be associated with an unusually high level of risk in our study (HR = 2.96 [1.97–4.46] for all-cause mortality, age-sex adjusted), which is more than twice the estimate from a recent meta-analysis (HR = 1.59 [1.42–1.77]) [3]. This strong effect may be plausible, however, if unemployment protections are weaker in Eastern Europe than elsewhere, which might highlight policy weaknesses for intervention.

Literature on material conditions has mostly focused on area-level measures of exposure [26, 27], while available studies at the individual level have often measured just 1–2 possessions, not aggregated such possessions into a summary score, or not controlled for blood pressure/cholesterol [28]. One comprehensive study in Russia did not find an association between material goods and/or amenities and all-cause mortality once education was controlled for [29]. Our larger study found the opposite pattern: that material amenities was the principal socioeconomic factor, which displaced education in multivariate analysis. Our analysis is consistent with prior reports of how education and material conditions may be measuring the same underlying socioeconomic construct, and it emphasises how material amenities might be a more sensitive socioeconomic predictor of cardiovascular mortality than education.

Putative mechanisms via conventional CVD risk factors

Previous studies have suggested that conventional risk factors might account for around 30%–50% of the primary association between socioeconomic factors and mortality [2, 30, 31], consistent with our analysis of education. However, in our data the proportion explained by conventional factors was much smaller for social support factors, such as only 3% for lack of contact with friends and 8% for single marital status. Therefore, the role of conventional risk factors might be even smaller among psychosocial factors (such as social support and also depression) that have been less studied to date.

Putative mechanisms via depression and social support

To our knowledge, this is the largest study to look at whether primary socioeconomic gradients in CVD can be attenuated following adjustment with psychosocial factors such as depression and social support. In Whitehall II, the occupational gradient in nonfatal coronary heart disease did not attenuate substantially following inclusion of social support measures [22], consistent with our findings. Instead, work-related stressors accounted for half of the occupational gradient in Whitehall II. We did not evaluate work factors given that a recent meta-analysis has shown their association with coronary heart disease to be comparatively small [24]. Other studies have typically used single-item instruments [32], or not assessed social support and conventional risk factors in the same study [30, 31]. Our study has confirmed relatively robustly that depression and social support do not account for much of socioeconomic gradients, even in a large cohort with high prevalence of exposure and outcome.

Our results suggest that the primary association between psychosocial factors (such as depression or social support) and CVD is unlikely to be mediated or confounded by conventional and other psychosocial risk factors, an area of limited prior study [33]. For example, elevated risk arising from exposure to one of the three dimensions of low social support was not offset by protection in another dimension, suggesting that cardiovascular health may be protected by some contact with friends, family, and a spouse. It appears that, in contrast to socioeconomic factors, each psychosocial factor is distinctly separate and does not relate to a common underlying construct. While attempts have been made to identify biomarkers of socioeconomic gradients, comparatively few have attempted to discover the mediators of psychosocial factors [23]. Our analysis suggests that depression is unlikely to be a major mediator of social support and socioeconomic pathways.

There is a paucity of studies investigating the potential mechanisms and mediators of psychosocial risk factors. We estimate, relatively crudely, that about one quarter of such hazards may be mediated by conventional CVD risk factors, and up to another quarter may come from other psychosocial factors such as depression. Measurement error in mediators biases our estimates of mediation towards the null, so future studies with time-varying mediators may be able to demonstrate larger proportions mediated [30]. In the case that such analyses fail to account entirely for the mechanistic pathways from psychosocial factors to CVD, additional hypotheses for investigations are warranted. These could include improved diet, or use of healthcare services (from better health knowledge, or fewer financial- and time-barriers in accessing care). Furthermore, direct molecular effects might be better clarified with emerging 'omics technologies.

International differences

We are not aware of other cohort studies that have used a standardised protocol to investigate why cardiovascular mortality is so dramatically high in Russia when directly compared to elsewhere. Ecological data from the Multinational MONItoring of trends and determinants in CArdiovascular disease (MONICA) study suggested that 0%–50% of the variations across time, in Europe and Russia, might be attributable to differences in conventional risk factors [7]. This is supported by studies of smoking and alcohol in Russia [34, 35]. Using different methods 30 years later, we find that conventional, psychosocial and socioeconomic factors do not account for much of the difference between Central European and Russian cohorts. We have previously shown that two other hypothesised factors, alcohol [15] and dietary factors [14], made only minor contributions to explaining the intercohort differences in mortality. After considering the results of the current study and the wider literature, we were unable to clarify the reasons for the very high cardiovascular mortality in Russian men. It is possible that international differences in access to healthcare, in healthcare seeking behaviour, or in the quality of care received, may help to clarify this question. Somewhat problematically, any such factors would need to have a large differential impact on male mortality but only a small differential impact on female mortality. Another area of enquiry might be gender norms and expectations. For example, women in Russia retire five years earlier than men. It is plausible that gender roles are more extremely polarised in such countries, which ultimately harms the health of men more than women [36].

Strengths and limitations

Several limitations of this study should be considered when interpreting the results. First, these urban population samples are not necessarily representative of whole countries, as both

exposures and mortality might be different in rural settings. Second, study participants may have been healthier than nonresponders, making us blind to what happens among those facing the greatest health and social problems. This would have led to underestimated HRs and population-attributable risk fractions. That said, our sample still detected considerable variation in cardiovascular mortality by country and socioeconomic status, as well as a considerable burden of psychological distress (e.g., 22% of participants screening positive for possible depression). Third, although we used predominantly well-established instruments to assess psychosocial exposures, their self-reported nature may be affected by response bias. For example, those who report adverse profiles might be more neurotic or less conscientious in their personality, which might instead be the underlying cause of our observed associations [37, 38]. Fourth, measurement error in mediators leads to underestimation in the degree of attenuation [39], and more importantly, any unmeasured confounders between mediators and outcome could have biased the results of our mediation analyses in either direction [40]. Fifth, our study makes no claims about causality due to its observational design. While some studies have supported a causal interpretation [41], the observational associations we report between CVD and social and psychosocial factors may be partly due to reverse causation or unmeasured confounding [42]. Sixth, it is uncertain how generalisable our results might be to other countries, given the social history and high mortality rates in this region.

As well as limitations, our study also has important strengths. First, this is the first prospective cohort study to our knowledge of psychosocial factors with a standardised methodology across multiple countries, where one country has twice the mortality rate of the others. Second, it is one of the largest prospective cohort studies that has assessed multiple psychosocial factors whilst concurrently controlling for all the conventional cardiovascular risk factors (including cholesterol and blood pressure). Third, most of the risk factors and covariates were measured using well-established and widely used psychosocial questionnaires and laboratory techniques.

Conclusion

Psychosocial and socioeconomic factors are powerful predictors of cardiovascular mortality in Eastern Europe, similarly to elsewhere. For some risk factors, such as unemployment, these associations appear stronger than elsewhere, indicating potential areas for policy intervention. Surprisingly, most of the associations between psychosocial risk factors and CVD were attenuated by only a small amount when adjusted for each other. This suggests that there may be a lot of nuanced pathways which could be relatively independent of one another, although we are aware that the causality of these associations requires further confirmation. The massive burden of cardiovascular mortality seen in Russia remains largely unexplained.

Supporting information

S1 Checklist. STROBE guidelines, for reporting cohort studies.
(DOCX)

S1 Table. Baseline data, stratified by missing status.
(DOCX)

S2 Table. Baseline data, stratified by country.
(DOCX)

S3 Table. Baseline data, stratified by gender.
(DOCX)

S4 Table. Conventional risk factors and cardiovascular mortality. 556 events among 20,867 participants.
(DOCX)

S5 Table. Psychosocial factors and cardiovascular mortality. Main analysis. 556 events among 20,867 participants.
(DOCX)

S6 Table. Psychosocial factors and cardiovascular mortality: gender interactions. Hazards greater than one indicate a higher hazard in male participants, compared to female participants.
(DOCX)

S7 Table. Psychosocial factors and cardiovascular mortality: country interactions. Hazards greater than one indicate a higher hazard in participants from Russia (versus Czech Republic/Poland).
(DOCX)

S8 Table. Psychosocial factors and cardiovascular mortality: complete cases. Data shows associations for 326 events, among those with no missing data on any covariate ($N = 13,727$).
(DOCX)

S9 Table. Psychosocial factors and cardiovascular mortality: excluding 2 years. Data shows associations for 458 events, among those with at least two years of follow-up ($N = 20,560$).
(DOCX)

S10 Table. Psychosocial factors and cardiovascular mortality: limiting follow-up to eight years. Data shows associations for 514 events, among all participants ($N = 20,867$), whereby follow-up time was censored at 8.0 years in all three countries.
(DOCX)

S11 Table. Psychosocial factors and all-cause mortality. 1,572 events among 20,867 participants.
(DOCX)

S12 Table. Attenuation of psychosocial factors. Hazard ratios for cardiovascular mortality (in rows) attenuate from left to right, following the sequential addition of psychosocial covariates (columns). 556 events among 20,867 participants.
(DOCX)

S13 Table. Attenuation of country-dummies. Hazard ratios for cardiovascular mortality from being in Russia (versus being in Central Europe, shown in rows) attenuates from left to right, following the sequential addition of psychosocial covariates (columns). 556 events among 20,867 participants. Results are stratified for men (380 events) and women (176 events).
(DOCX)

Author Contributions

Conceptualization: Taavi Tillmann, Martin Bobak.

Data curation: Hynek Pikhart, Anne Peasey, Ruzena Kubinova, Andrzej Pajak, Abdonas Tamosiunas, Sofia Malyutina.

Formal analysis: Taavi Tillmann.

Funding acquisition: Taavi Tillmann, Sofia Malyutina, Martin Bobak.

Methodology: Hynek Pikhart, Andrew Steptoe, Mika Kivimäki, Martin Bobak.

Project administration: Hynek Pikhart, Anne Peasey, Martin Bobak.

Supervision: Michael Marmot, Martin Bobak.

Visualization: Taavi Tillmann.

Writing – original draft: Taavi Tillmann.

Writing – review & editing: Taavi Tillmann, Hynek Pikhart, Anne Peasey, Ruzena Kubinova, Andrzej Pajak, Abdonas Tamosiunas, Sofia Malyutina, Andrew Steptoe, Mika Kivimäki, Michael Marmot, Martin Bobak.

References

1. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. American Journal of Epidemiology. 1979; 109(2):186–204. PMID: 425958
2. Veronesi G, Ferrario MM, Kuulasmaa K, Bobak M, Chambless LE, Salomaa V, et al. Educational class inequalities in the incidence of coronary heart disease in Europe. Heart. 2016; 102(12):958–65. https://doi.org/10.1136/heartjnl-2015-308909 PMID: 26849899
3. Roelfs DJ, Shor E, Davidson KW, Schwartz JE. Losing life and livelihood: a systematic review and meta-analysis of unemployment and all-cause mortality. Social Science & Medicine. 2011; 72(6):840–54. https://doi.org/10.1016/j.socscimed.2011.01.005 PMID: 21330002
4. Barth J, Schneider S, von Känel R. Lack of social support in the etiology and the prognosis of coronary heart disease: a systematic review and meta-analysis. Psychosomatic Medicine. 2010; 72(3):229–38. https://doi.org/10.1097/PSY.0b013e3181d01611 PMID: 20223926
5. Van der Kooy K, van Hout H, Marwijk H, Marten H, Stehouwer C, Beekman A. Depression and the risk for cardiovascular diseases: systematic review and meta analysis. International Journal of Geriatric Psychiatry. 2007; 22(7):613–26. https://doi.org/10.1002/gps.1723 PMID: 17236251
6. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). European Heart Journal. 2016; 37(29):2315–81. https://doi.org/10.1093/eurheartj/ehw106 PMID: 27222591
7. Kuulasmaa K, Tunstall-Pedoe H, Dobson A, Fortmann S, Sans S, Tolonen H, et al. Estimation of contribution of changes in classic risk factors to trends in coronary-event rates across the WHO MONICA Project populations. The Lancet. 2000; 355(9205):675–87. http://dx.doi.org/10.1016/S0140-6736(99)11180-2 PMID: 10703799
8. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2017 [Accessed on 20 sept 2017]. Available from: https://esa.un.org/unpd/wpp/DataQuery/.
9. Wang H, Naghavi M, Allen C, Barber RM, Zulfiquar AB, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016; 388(10053):1459–544. http://dx.doi.org/10.1016/S0140-6736(16)31012-1 PMID: 27733281
10. Bobak M, Marmot M. East-West mortality divide and its potential explanations: proposed research agenda. BMJ. 1996; 312:421. https://doi.org/10.1136/bmj.312.7028.421 PMID: 8601115
11. Hajdu P, Mckee M, Bojan F. Changes in premature mortality differentials by marital status in Hungary and in England and Wales. The European Journal of Public Health. 1995; 5(4):259–64. https://doi.org/10.1093/eurpub/5.4.259
12. Shkolnikov VM, Leon DA, Adamets S, Andreev E, Deev A. Educational level and adult mortality in Russia: an analysis of routine data 1979 to 1994. Social Science & Medicine. 1998; 47(3):357–69. https://doi.org/10.1016/S0277-9536(98)00096-3 PMID: 9681906
13. Peasey A, Bobak M, Kubinova R, Malyutina S, Pajak A, Tamosiunas A, et al. Determinants of cardiovascular disease and other non-communicable diseases in Central and Eastern Europe: rationale and design of the HAPIEE study. BMC Public Health. 2006; 6:255. https://doi.org/10.1186/1471-2458-6-255 PMID: 17049075

14. Stefler D, Pikhart H, Jankovic N, Kubinova R, Pajak A, Malyutina S, et al. Healthy diet indicator and mortality in Eastern European populations: prospective evidence from the HAPIEE cohort. European Journal of Clinical Nutrition. 2014; 68(12):1346. https://doi.org/10.1038/ejcn.2014.134 PMID: 25028084

15. Bobak M, Malyutina S, Horvat P, Pajak A, Tamosiunas A, Kubinova R, et al. Alcohol, drinking pattern and all-cause, cardiovascular and alcohol-related mortality in Eastern Europe. European Journal of Epidemiology. 2016; 31(1):21–30. https://doi.org/10.1007/s10654-015-0092-8 PMID: 26467937

16. Vandenheede H, Vikhireva O, Pikhart H, Kubinova R, Malyutina S, Pajak A, et al. Socioeconomic inequalities in all-cause mortality in the Czech Republic, Russia, Poland and Lithuania in the 2000s: findings from the HAPIEE Study. Journal of Epidemiology and Community Health. 2014; 68(4):297–303. https://doi.org/10.1136/jech-2013-203057 PMID: 24227051

17. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement. 1977; 1(3):385–401. https://doi.org/10.1177/014662167700100306

18. Bobak M, Pikhart H, Hertzman C, Rose R, Marmot M. Socioeconomic factors, perceived control and self-reported health in Russia. A cross-sectional survey. Social Science & Medicine. 1998; 47(2):269–79. https://doi.org/10.1016/S0277-9536(98)00095-1 PMID: 9720645

19. Collaboration PS. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. The Lancet. 2002; 360(9349):1903–13. http://dx.doi.org/10.1016/S0140-6736(02)11911-8 PMID: 12493255

20. Rehm J. Measuring quantity, frequency, and volume of drinking. Alcoholism: Clinical and Experimental Research. 1998; 22(S2). PMID: 9603301

21. Lane PW, Nelder JA. Analysis of covariance and standardization as instances of prediction. Biometrics. 1982:613–21. PMID: 7171691

22. Marmot MG, Bosma H, Hemingway H, Brunner E, Stansfeld S. Contribution of job control and other risk factors to social variations in coronary heart disease incidence. The Lancet. 1997; 350(9073):235–9. http://dx.doi.org/10.1016/S0140-6736(97)04244-X PMID: 9242799

23. Nabi H, Singh-Manoux A, Shipley M, Gimeno D, Marmot MG, Kivimaki M. Do psychological factors affect inflammation and incident coronary heart disease. Arteriosclerosis, Thrombosis, and Vascular Biology. 2008; 28(7):1398–406. https://doi.org/10.1161/ATVBAHA.108.167239 PMID: 18436803

24. Kivimäki M, Nyberg ST, Batty GD, Fransson EI, Heikkilä K, Alfredsson L, et al. Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data. The Lancet. 2012; 380(9852):1491–7. http://dx.doi.org/10.1016/S0140-6736(12)60994-5 PMID: 22981903

25. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for individual Prognosis or diagnosis (TRIPOD): Explanation and Elaboration. Annals of Internal Medicine. 2015; 162(1):W1–W73. https://doi.org/10.7326/M14-0698 PMID: 25560730

26. Filakti H, Fox J. Differences in mortality by housing tenure and by car access from the OPCS Longitudinal Study. Population Trends. 1995;(81):27–30. PMID: 8528790

27. Macleod J, Smith GD, Metcalfe C, Hart C. Is subjective social status a more important determinant of health than objective social status? Evidence from a prospective observational study of Scottish men. Social Science & Medicine. 2005; 61(9):1916–29. https://doi.org/10.1016/j.socscimed.2005.04.009 PMID: 15916842

28. van Oort FV, van Lenthe FJ, Mackenbach JP. Material, psychosocial, and behavioural factors in the explanation of educational inequalities in mortality in The Netherlands. Journal of Epidemiology and Community Health. 2005; 59(3):214–20. https://doi.org/10.1136/jech.2003.016493 PMID: 15710599

29. Perlman F, Bobak M. Socioeconomic and behavioral determinants of mortality in posttransition Russia: a prospective population study. Annals of Epidemiology. 2008; 18(2):92–100. https://doi.org/10.1016/j.annepidem.2007.07.093 PMID: 17923417

30. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimaki M, et al. Association of socioeconomic position with health behaviors and mortality. JAMA. 2010; 303(12):1159–66. https://doi.org/10.1001/jama.2010.297 PMID: 20332401

31. Kershaw KN, Droomers M, Robinson WR, Carnethon MR, Daviglus ML, Verschuren WM. Quantifying the contributions of behavioral and biological risk factors to socioeconomic disparities in coronary heart disease incidence: the MORGEN study. European Journal of Epidemiology. 2013; 28(10):807–14. https://doi.org/10.1007/s10654-013-9847-2 PMID: 24037117

32. Fiscella K, Franks P. Does psychological distress contribute to racial and socioeconomic disparities in mortality? Social Science & Medicine. 1997; 45(12):1805–9. https://doi.org/10.1016/S0277-9536(97)00111-1 PMID: 9447630

33. Hintsa T, Shipley MJ, Gimeno D, Elovainio M, Chandola T, Jokela M, et al. Do pre-employment influences explain the association between psychosocial factors at work and coronary heart disease? The Whitehall II study. Occupational and Environmental Medicine. 2010; 67(5):330–4. https://doi.org/10.1136/oem.2009.048470 PMID: 19819857

34. Leon DA, Saburova L, Tomkins S, Andreev E, Kiryanov N, McKee M, et al. Hazardous alcohol drinking and premature mortality in Russia: a population based case-control study. The Lancet. 2007; 369 (9578):2001–9. https://doi.org/10.1016/S0140-6736(07)60941-6 PMID: 17574092

35. Zaridze D, Lewington S, Boroda A, Scélo G, Karpov R, Lazarev A, et al. Alcohol and mortality in Russia: prospective observational study of 151 000 adults. The Lancet. 2014; 383(9927):1465–73. https://doi.org/10.1016/S0140-6736(13)62247-3 PMID: 24486187

36. Watson P. Explaining rising mortality among men in Eastern Europe. Social Science & Medicine. 1995; 41(7):923–34. https://doi.org/10.1016/0277-9536(94)00405-I PMID: 8545667

37. Jokela M, Batty GD, Nyberg ST, Virtanen M, Nabi H, Singh-Manoux A, et al. Personality and all-cause mortality: individual-participant meta-analysis of 3,947 deaths in 76,150 adults. American Journal of Epidemiology. 2013; 178(5):667–75. https://doi.org/10.1093/aje/kwt170 PMID: 23911610

38. Jokela M, Pulkki-Råback L, Elovainio M, Kivimäki M. Personality traits as risk factors for stroke and coronary heart disease mortality: pooled analysis of three cohort studies. Journal of Behavioral Medicine. 2014; 37(5):881–9. https://doi.org/10.1007/s10865-013-9548-z PMID: 24203126

39. Blakely T, McKenzie S, Carter K. Misclassification of the mediator matters when estimating indirect effects. Journal of Epidemiology and Community Health. 2013; 67(5):458–66. https://doi.org/10.1136/jech-2012-201813 PMID: 23386673

40. Hafeman DM. Confounding of indirect effects: a sensitivity analysis exploring the range of bias due to a cause common to both the mediator and the outcome. American Journal of Eepidemiology. 2011; 174 (6):710–7. https://doi.org/10.1093/aje/kwr173 PMID: 21652602

41. Tillmann T, Vaucher J, Okbay A, Pikhart H, Peasey A, Kubinova R, et al. Education and coronary heart disease: mendelian randomisation study. BMJ. 2017; 358:j3542. https://doi.org/10.1136/bmj.j3542 PMID: 28855160

42. Martikainen PT, Valkonen T. Excess mortality of unemployed men and women during a period of rapidly increasing unemployment. The Lancet. 1996; 348(9032):909–12. https://doi.org/10.1016/S0140-6736(96)03291-6 PMID: 8843808